

Anti-terrorist agendas

The current security crisis will lead to the restoration of an intimate relationship between science and the US federal government, in which money-grubbing will take a back seat.

The events of 11 September have shaken many people to the core, scientists included, and it is naturally taking time for their various practical implications to sink in. But one thing is clear: in time of war, modern governments require a special relationship with their best scientists and engineers — a relationship that now needs to be recast.

In the United States, such a relationship was forged by President Franklin Roosevelt and Vannevar Bush, the first real White House science adviser, during the Second World War, and prospered during the early years of the cold war. In the decade of American complacency that followed the cold war's end, it fell into considerable disrepair.

Under President Bill Clinton, the President's Council of Advisors on Science and Technology (PCAST), the White House Office of Science and Technology Policy (OSTP) and the National Science and Technology Council sought to supervise the massive research and development exercise marshalled by the US federal government. The OSTP made some successful forays — into the sharing of satellite data between military and civilian agencies, for example — but its reach was minimal. It is a measure of the limited influence of these bodies that all have lain dormant this past year, with no obvious consequences.

That should change now. At a meeting at the National Academy of Sciences on 26 September, an array of heavyweights, including John Marburger, President George W. Bush's nominated science adviser, Bruce Alberts, the president of the National Academy of Sciences, Sig Hecker, Joshua Lederberg, Sam Nunn, James Woolsey, Wolfgang Panofsky, Richard Garwin, Maxine Singer and Richard Klausner, met privately to discuss the form that the change should take.

These luminaries don't have all the answers, and would not pretend to do so. But they do have some knowledge of the relationship

between science and government. One of their first proposals was a new approach to research related to counter-terrorism. Naturally, preparedness for bioterrorism is part of that — even on 26 September, few could have foreseen how quickly it would rear its ugly head.

But the scope of the challenge is unprecedented. It is no longer possible for a few well-placed advisers to grasp all of the disciplines and subdisciplines of the United States' \$250-billion-a-year science and technology enterprise (\$90 billion in the public sector, the rest in industry). Expertise in some of the most important fields, including biotechnology and computer software, is widely dispersed, much of it outside the purview of government.

Federal advisory panels such as PCAST will have an important role in establishing links between civilian researchers and the government. But these panels will need to be reconstituted to function with an urgency and flexibility that they previously lacked.

Furthermore, the basic dynamics of relations between scientific leaders and the government will have to change. For too long, the central characteristic of these relations has been the scientists' unending quest for more money. Already, the community's discussions with the Congress have focused unduly on the need for more money for intelligence, or more money to counter bioterrorism.

The community needs to get past this. In the coming months, a long, hard look is likely to be taken at the shape of the US federal government's hefty investment in research and development. It is by no means clear that all existing activities will be sustainable at a time when the government will be required to focus much of its attention on America's national-security crisis. Scientists are going to be a part of this reorientation, perhaps in far larger numbers and for a longer period of time than most of us care to imagine. ■

Top-heavy and out of touch?

A powerful new agency needs to attain a more appropriate balance between Japan's policy-makers and researchers.

If there's one thing that almost all Japanese scientists, science policy-makers and industrialists agree on, it is that Japan would benefit from more coordination between its various governmental science activities. To that end, the cabinet-level Council for Science and Technology Policy (CSTP) was born earlier this year.

Some claim that the CSTP has already overstepped its bounds in its first year by making specific budget decisions rather than merely defining broad national priorities. Others would like to give it more power, for example to distribute grants, and there is even discussion about creating an integrated biosciences body — what some say would be a Japanese National Institutes of Health (NIH) — under its auspices (see page 659).

Concerns that the CSTP is too bureaucratic — ministers dominate the CSTP in numbers and influence — are mirrored by fears that any NIH-like organization would be completely 'top-down' in operation. The worst-case scenario is that these non-scientists would back only the kind of science that produces short-term gain and expands their

own ministries' budgets. Many researchers feel alienated from and distrustful of this body, which should be representing them in the government. They plead that a certain amount, say 60%, of spending on research and development should be set aside for basic science.

No integrated science effort will be successful in Japan without a greater voice from scientists in the government. The CSTP should include more than the one active scientist it has now. The government must learn to broaden its group of decision-makers. If implemented correctly, the CSTP could come into its own over decisions about setting up an NIH-like organization or by taking a leading role in supporting such a body. Otherwise, the worst fears of Japanese scientists could be realized.

For their part, researchers must make their voices heard. Too often they wait for their opinions to be asked, and complain when they are not, but at the same time view such political activity as getting their hands dirty. They should, rather, actively voice their concerns to their institute directors, their politicians and, not least, the CSTP. ■





Unprepared
Japan still vulnerable
to gas and bioweapon
attacks
p658



Not flying high
Farm biodiversity
schemes fail to raise
bird populations
p659



Costly strategy
CERN director
general denies poor
budget management
p660



Mirror molecules
Nobel chemistry
award has a pleasing
symmetry
p661

Genetic sleuths rush to identify anthrax strains in mail attacks

Rex Dalton, San Diego

As reports of anthrax attacks across the United States multiplied earlier this week, researchers scrambled to identify the strains of bacteria involved. Such information will be vital in shaping the investigation of the attacks, and the political response to them.

Investigators hope that new analytical approaches will shed light on the origins of anthrax spores sent through the mail. As *Nature* went to press, envelopes containing spores had already been confirmed in Florida, New York, Nevada and at the offices of majority senate leader Tom Daschle (Democrat, South Dakota) in Washington. If they all involve the same strain, for example, it would indicate that they are part of a coordinated attack.

As the number of reported attacks rose over the weekend, health secretary Tommy Thompson said that the White House would ask Congress for an additional \$1.5 billion to counter bioterrorism and purchase antibiotics.



Handled with care: FBI agents remove suspicious material from newspaper offices in Florida.

According to sources close to the investigation, the laboratory of Paul Keim, a geneticist at the Northern Arizona University (NAU) in Flagstaff, has taken an early lead in identifying

the strains of the causative bacterium, *Bacillus anthracis*, involved in the attacks.

The authorities turned to Keim's lab immediately after the Florida attacks because of its published record of describing anthrax strains and its role in developing analytical techniques to identify the strains.

In 1998, Keim's team played a key role in proving that a deadly outbreak of anthrax at Sverdlovsk in Russia involved multiple strains of the bacterium, and so was related to an escape of material from germ-warfare facilities and not a natural outbreak, as Russia had claimed (P. J. Jackson *et al. Proc. Natl Acad. Sci. USA* **95**, 1224–1229; 1998).

Although Keim declined to discuss specifics of the latest investigation or the results of his tests, authorities have said the anthrax spores in the Florida attack are believed to be a derivative of the virulent 'Ames strain'. This was commonly used in laboratories for research, developing vaccines and tests, after its original isolate was removed from a dead animal in the 1950s near Ames, Iowa.

More than 1,200 strains of *B. anthracis* have been identified around the world over the years, and the NAU laboratory has used AFLP (amplified fragment length polymorphism) DNA analysis to examine all of them.

This technique uses enzymes to cut the

Bioweapons treaty under threat

Declan Butler, Paris

Even as envelopes containing anthrax spores are increasing fears of bioterrorism in the United States, a flashpoint looms for the international treaty intended to help prevent such attacks.

A lack of consensus over verification threatens to unravel the 1972 Biological Weapons Convention (BWC) at a meeting next month.

Years of efforts to give the BWC teeth collapsed in July, when the United States vetoed a proposed verification protocol for the treaty (see *Nature* **412**, 365; 2001).

US officials last week confirmed that the 11 September attacks had not shifted their position. Avis Bohlen, assistant secretary of state for arms control, told a United Nations forum that the events had "reinforced our view" that verification was unworkable.

More detailed proposals of verification

measures were put forward in Paris last week by US officials at a closed meeting of the Australia Group, an informal alliance of 33 countries that aims to limit the proliferation of dangerous agents and technologies through export controls. But other delegates were unconvinced that the United States can suggest a meaningful alternative to the protocol it rejected in July.

The next of the BWC's five-yearly review meetings is scheduled for next month, and the treaty requires all participants to reach a consensus, however bland, at such meetings. Officials fear that some parties, such as Iran or Cuba, may attack the US verification veto, causing a complete breakdown in the process.

"If you don't get consensus, then you don't have another conference in five years' time, and then the BWC is dead," says Julian Perry Robinson, a disarmament expert at the University of Sussex, UK. ■

bacterium's genome into random fragments, and then uses the polymerase chain reaction (PCR) to amplify some of the fragments. Sequence variation between strains, and variation in the length of regions in which small stretches of sequence are repeated, results in a different pattern of amplified fragments for each strain. Data derived from the AFLP technique are held on databases at the NAU and at the Los Alamos National Laboratory in New Mexico.



Taking the strain:
Paul Keim.

But Keim's lab has also adapted a more precise test called multi-locus VNTR (variable-number tandem repeat) analysis, or MLVA, for use with microorganisms. With MLVA, researchers use PCR to amplify regions of the genome containing repeated sequences and so develop a genetic fingerprint for a strain or species. MLVA has so far been performed on about 400 of the 1,200 known strains of *B. anthracis*.

The genome of *B. anthracis* contains short sequences of DNA that, depending on the strain, are repeated a different number of times at certain loci. Keim's group has published findings on eight such markers used to probe these loci, using a technique that is similar to that used in genetic fingerprinting for paternity suits or criminal investigations. Keim's team has a total of about 50 *B. anthracis* markers at its disposal, and is working on more markers for the roughly 1,000 VNTR loci in the bacterium's genome. "This is a highly precise method that also can be used for tuberculosis, *Escherichia coli* and other pathogens," says Keim.

It takes about 12 hours for Keim's lab to analyse an anthrax sample. Once a strain is identified, investigators can try to match it to its original source — but this is not always easy. The Ames strain, for example, has been passed around the world by researchers — the sample held by the NAU came from Britain's chemical and biological defence facility at Porton Down, which received it from the US Army Medical Research Institute of Infectious Diseases at Fort Detrick in Maryland.

In the future, analysis of anthrax will get a boost from the sequencing of a derivative of the Ames strain by The Institute for Genomic Research (TIGR) in Maryland. Timothy Read, a bacterial genomicist at TIGR, says he expects to close the final gaps in the strain's sequence in the next few months. Other strains will also be sequenced in the near future.

▶ <http://herb.bio.nau.edu/~genetics/project3.htm>

Scientific leaders respond to US government's call to arms

Jonathan Knight, San Francisco

Leaders of the US scientific community are developing plans to work closely with government agencies and the military in the fight against terrorism.

On 26 September, for example, Bruce Alberts, president of the National Academy of Sciences, convened a meeting of 30 prominent scientists and security experts to discuss the role of science in the aftermath of the 11 September attacks on New York and Washington.

Those attending included Norman Augustine, chairman of aeronautics company Lockheed Martin; John Marburger, director of the Department of Energy's Brookhaven National Laboratory and President George W. Bush's nominated science adviser; Wolfgang Panofsky, a director of the Stanford Linear Accelerator Center; former senator Sam Nunn; and James Woolsey, former head of the Central Intelligence Agency.

"All were clearly concerned that the scientific and technical community hasn't played an important role in the issue of dealing with organized terrorism," says Kumar Patel, a former president of the American Physical Society, who also attended the meeting.

Several federal agencies, including the National Science Foundation and the Defense Advanced Research Projects Agency, currently fund counter-terrorism research. But according to Maxine Singer, president of the Carnegie Institution, who was present at the meeting, they sometimes fail to draw on the full creative potential of the academic community. "Things have become institutionalized," she says.

Both the White House Office of Science and Technology Policy and the Pentagon's Technical Support Working Group have approached the academy complex to ask for input, says Bill Colglazier, executive director of the National Research Council. "They want cutting-edge people to help with brainstorming and advising," he says.

To coordinate this input, the academies are planning to spend \$500,000 of their own money to help create a task force with an anti-terrorist research agenda, called the Multi-Agency Program Plan for Science and Technology. Lewis Branscomb, a technology-policy specialist at Harvard University, and Richard Klausner, a former director of the National Cancer Institute, are leading the project.

Gaps remain in Japan's biodefences

David Cyranoski, Tokyo

Japan remains vulnerable to bioterrorist and chemical attacks, citizen groups and biomedical researchers are warning. Despite

suffering the 1995 sarin nerve-gas attack on the Tokyo subway, the government has not taken sufficient action, the groups say.

"Now many hospitals and clinics are aware of emergency measures for sarin treatment," says Masanori Fukushima, an epidemiologist at Kyoto University, "but there is no preparation for recognizing victims of bioterrorism."

Yukitatsu Kawamoto, of the Citizen's Centre for the Prevention of Biohazards in Chiba, notes that in 1993 Aum Shinrikyo, the cult behind the sarin attack, sprayed a non-virulent strain of anthrax from the roof of its Tokyo headquarters. "The authorities never tracked down where Aum got it, and they haven't taken measures to account for such materials since then," he says.

And last week Yasuo Fukuda, the government's chief cabinet secretary, admitted that Japan's anti-terrorism preparations are inadequate.

But Toshinobu Sato, director of the Office of Health Crisis Management, claims that security at health-ministry institutes is sufficient to prevent anyone from removing potential biological weapons.



History unheeded? Despite 1995's sarin gas attack, Japan is ill-prepared to fight terrorism.

Birds fly in the face of 'green' farming incentive scheme

John Whitfield, London

The effectiveness of schemes that seek to promote biodiversity by paying farmers to cut back on intensive agriculture could be called into question by some research findings from Holland.

The incentive programmes, which already cost the European Union 1.7 billion euros (US\$1.5 billion) each year and are rapidly expanding in scope, are partly motivated by the desire of European governments to subsidize farming without promoting the overproduction of food. But the programmes have been publicly justified by claims that they will help to restore ecological biodiversity on farmland.

In this issue of *Nature*, David Kleijn and colleagues at Wageningen University report results from one of the first schemes of this kind, operated by the Dutch government since 1981. They find that it has failed to significantly increase biodiversity in fields where farmers were paid to delay mowing or grazing, and to reduce the amount of fertilizer they used (see page 723).

The study did not examine how the total biodiversity in the region had changed over time, however. And EU officials and some conservationists are confident that the more sophisticated programmes now being implemented will increase biodiversity.

But many ecologists claim that such programmes have not been properly evaluated. "There are far too few studies of this kind," says population biologist Arie van Noordwijk of the Netherlands Institute of Ecology's Centre for Terrestrial Ecology in Heteren. "We have hardly any good data on the effectiveness of these programmes, and vast amounts of money are spent on them Europe-wide."

Kleijn's study counted species and numbers of plants, birds, bees and hoverflies in 78 fields last summer. Only bee and hoverfly



Dutch dilemma: the black-tailed godwit (left) is among birds that have not benefited from a scheme to limit intensive agriculture.

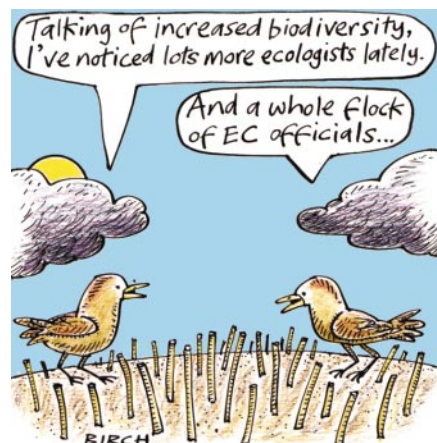


diversity increased under the scheme. Birds actually seemed to prefer intensively farmed fields, possibly because reductions in fertilizer use lead to smaller invertebrate populations and so less food.

Kleijn thinks the study area is representative of intensively managed farmland across northern Europe, but admits that the effectiveness of such schemes in hill-farm regions, for example, may be quite different.

According to Carlos Martin-Novella, an official at the European Commission's Development and Environment Unit in Brussels, the European Union has already adjusted its subsidy schemes to increase their effectiveness, and made arrangements to monitor them. For example, the commission's biodiversity action plan, published earlier this year, requires member states to evaluate their agri-environment schemes. The new research "fits very well into this commitment", says Martin-Novella.

David Gibbons, head of conservation science at the Royal Society for the Protection of Birds, a British conservation group, says that recent environmental subsidy schemes are designed with more scientific input than the Dutch study. Britain's Arable Stewardship project, for example, which aims to restore farmland bird populations, expressly targets individual species. It will require farmers to leave stubble — an important food source for birds — in fields over the winter, instead of digging it up to plant crops such as winter wheat. The scheme is to be launched in 2002 after a three-year pilot study. "I'd be shocked if we don't see real benefits over the next 5–10 years," says Gibbons.



Plans to centralize biology research set for rough ride

David Cyranoski, Tokyo

A special summit meeting in Tokyo next week is expected to be dominated by arguments over whether Japan needs a new life-sciences research agency akin to the US National Institutes of Health (NIH).

The Japan Association of Bioindustries Executives (JABEX), one of the groups hosting the meeting, plans to use it to publicly unveil an ambitious plan to streamline support mechanisms for life-sciences research in Japan. Currently, such support comes from several different sources, including the health, education and economics ministries.

Although details of the proposal remain sketchy, it is known to involve the creation of a powerful biomedical research agency with responsibility for both basic and clinical research. The new body would seek to bolster Japan's flagging clinical-research efforts and encourage more rapid commercialization of research findings.

But some Japanese researchers are worried that such a centralized body might stifle their independence. They are also unhappy that JABEX has taken its proposal directly to the government's Council for Science and Technology Policy (CSTP), instead of asking them first.

"They are approaching the CSTP without first consulting with researchers, and this makes me uneasy," says Takeshi Gojobori, head of the bioinformatics unit at the National Institute of Genetics in Mishima.

Osamu Chisaki, secretary of JABEX, says the CSTP would play a central role in implementing the plan.

The idea has already come under fire from officials at the education ministry, which currently supports most basic biology research. They claim that an NIH-type agency wouldn't work in Japan because it would become too bureaucratic. Research-policy decisions would become top-down and basic research would suffer, predicts one University of Tokyo researcher.

The summit, to be held on 22 October, will bring academics, such as Nobel laureate Susumu Tonegawa of the Massachusetts Institute of Technology, together with politicians, government officials and industry representatives. They will discuss the possibilities for restructuring biomedical research in Japan in line with the government reform efforts of Prime Minister Junichiro Koizumi, who will afterwards attend a reception.

CERN's head rejects mismanagement claims

Alison Abbott, Munich

Luciano Maiani, director general of CERN, the European particle physics laboratory in Geneva, has denied that mismanagement has led to cost overruns on the Large Hadron Collider (LHC) project.

CERN shocked its 20 member states last month by announcing cost overruns of SFr480 million (US\$296 million) on the SFr2.6-billion project, and of a further SFr370 million on CERN's core budget during the LHC's construction period (see *Nature* **413**, 441; 2001).

"We've always been in control of developments," says Maiani. "This is a high-tech project of extreme complexity, and an overrun of 18% is not unreasonable."

But Maiani, who was head of the Italian Nuclear Physics Institute before taking up his position at CERN in 1999, accepts the charge that his management team failed to reveal the scale of the problem as soon as it became apparent in the spring. "We've learnt a hard lesson about the need for openness," he says.

The problems should have been identified last year, when the first cost-to-completion estimates for the LHC were due, Maiani admits. But he says that he was distracted at the time by preliminary results suggesting that detectors at the Large Electron-Positron (LEP) collider, which the LHC will replace, had picked up signatures of the Higgs boson — the fundamental particle that the LHC is designed to find.

CERN researchers were then torn between extending the LEP's life in a bid to confirm the finding, or switching it off to

start work on the LHC. Maiani decided to pull the plug on the LEP — rightly as it turned out, as reanalysis of the preliminary results indicated little likelihood that the Higgs particle had actually been spotted. "The LEP decision fully absorbed the energy

of both the scientists and the administration," says Maiani.

A major project review was begun in March of this year, and final quotes for the supply of the LHC's 1,000 or so superconducting magnets — which Maiani decided to await before completing the review — were received from contractors in August.

"If I had the time again, I would have discussed the emerging financial difficulties with our scientists and with member-state delegations," he says. "But I decided to wait until we had solid numbers to discuss. This was probably the wrong decision — it backfired on us."

More unreasonable than the cost overrun, in Maiani's view, is the fact that the LHC project had to be launched on a tight budget with no financial contingency.

Asked how he will now proceed, Maiani warns of austerity ahead. "There's no doubt that we will have to streamline all of CERN's activities behind the LHC," he says. Smaller programmes in heavy-ion, neutrino and antimatter research will have to be cut back, and other laboratory costs will be reduced. He also suggests that loans could be sought to buy time — and hopes that member states will increase their contributions.

Maiani says he is confident that the project will get back on track, and even claims that last year's decision to turn off the LEP was much more difficult for him than resolving the new financial difficulties, which he insists are inherent in high-technology projects such as this. The technological progress of the LHC remains "rock-solid", he says. ■



A hard lesson learnt: Maiani accepts that delaying admission of financial problems was a mistake.

Sanger Centre welcomes gene funds with a new name

David Adam, London

The Sanger Centre near Cambridge, UK, has announced a five-year plan to spend £300 million (US\$435 million) on research to follow up on its role in the publication, earlier this year, of the human genome sequence (see *Nature* **409**, 860–923; 2001).

The plan, which will be paid for by the Wellcome Trust, the world's largest biomedical research charity, includes increased efforts to understand gene function and expression, to develop bioinformatics tools and to draw comparisons between humans and model organisms such as the mouse. The centre is also changing its name to become the Wellcome Trust Sanger Institute.

"The new funding will allow the institute to make a contribution to global science and

medicine as significant as its contribution to the Human Genome Project," says institute director Allan Bradley. "We will bring biology to the genome and translate the enormous amount of information encoded in our DNA into an understanding of gene function."

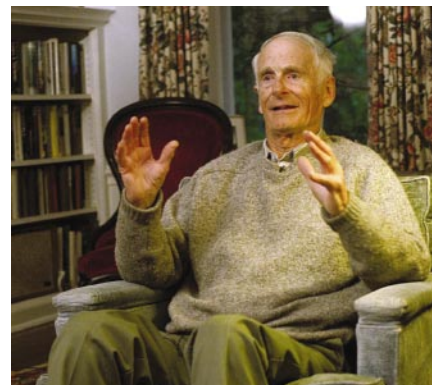
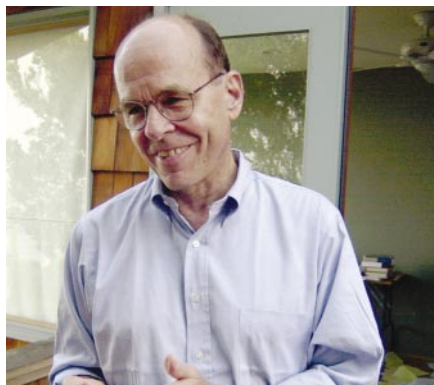
According to the new plan, the institute will work to identify as many genes as possible within the human genome and to predict their functions, as well as locating the regions that control their expression. The laboratory will also target disease genes, including those that contribute to cancers, building on a major project already under way to search for the genetic mutations that cause them.

One new project aims to identify genes on the X chromosome that are involved in primary or non-specific X-linked mental

retardation, one of the most common genetic disorders. Another will focus on uncovering the genetic basis of disorders that have multiple genetic causes, such as diabetes, asthma and other allergies. This association study will use a map of 'haplotype' groups — inherited blocks of genetic markers — to be prepared by an international consortium of labs and biotech companies (see *Nature* **412**, 105; 2001).

The institute will also establish a genetics programme to further the understanding of the role of genes in mouse development and physiology, and in human disease. "These new initiatives will complement and build on the institute's considerable strengths in genome data generation, automation and bioinformatics," says Mike Dexter, director of the Wellcome Trust. ■

Chemistry prize reflects tailor-made reactions



Masters of the mirror image: this year's chemistry laureates, from left to right, Barry Sharpless, Ryoji Noyori and William Knowles.

David Adam, London

The ability to steer chemical reactions towards one of two possible symmetrical products is important for the industrial synthesis of many drugs. For three researchers behind the technique of 'catalytic asymmetric synthesis', it also means a Nobel prize for chemistry.

Half of this year's prize goes to K. Barry Sharpless of the Scripps Research Institute in La Jolla, California, for his work on oxidation reactions. The other half is shared by two researchers for their work on hydrogenation reactions: Ryoji Noyori of Nagoya University, Japan, and William

Knowles, a retired chemist who worked for Monsanto in St Louis, Missouri, until 1986.

The winners studied reactions that, under normal circumstances, produce an equal mixture of two products with symmetrical structures. Known as chiral products, these molecules are mirror images of each other, but can have very different properties. The side-effects of the morning-sickness drug thalidomide, for example, were caused by a rogue chiral twin of the molecule that the drug's designers had intended to use.

Sharpless, Noyori and Knowles have been awarded the prize for developing techniques that tailor reactions so that only one of

the two chiral molecules is produced. The techniques are now widely used in industry, particularly to make pure pharmaceuticals.

In 1968 Knowles discovered that, by using a chiral catalyst, he could tweak a hydrogenation reaction, in which hydrogen is added to organic molecules, so that it produced 15% more of one chiral product. His team went on to develop a method for industrial synthesis of the amino acid L-DOPA (3,4-dihydroxyphenylalanine), a chiral molecule that is used to treat Parkinson's disease.

Knowles's work was extended by Noyori, who developed further catalysts, some of which can hydrogenate many types of molecule. Reactions using his catalysts give very pure products and high yields, and are now widely used to manufacture large quantities of antibiotics.

Noyori's success is the second by a Japanese chemist in as many years — taking Japan a little way towards its government's stated, and occasionally ridiculed, objective of obtaining 30 Nobel prizes in the next 50 years (see *Nature* **413**, 560–564; 2001).

"I'd be very happy if this Nobel could be useful in promoting basic science in Japan," says Noyori. "Japanese have won many international prizes in chemistry, and I'm happy that this strength is being acknowledged."

Sharpless developed chiral catalysts for oxidation reactions during the 1980s. One of these, which allows the asymmetric oxidation of a class of alcohols, is seen by many chemists as one of the most important discoveries made in the field of synthesis during the past few decades. One product, glycidol, is now used in the production of beta-blockers.

But some researchers are disappointed that Henri Kagan, a chemist at Paris-Sud University who also worked on asymmetric synthesis, missed out on the prize. Together with Sharpless and Noyori, Kagan won this year's Wolf Foundation Prize in chemistry, which is sometimes regarded as an indicator of future Nobel success.

Bid to end EU's transgenic impasse

Quirin Schiermeier, Munich

The European Commission is preparing to restart the approval process for commercial planting of genetically modified (GM) crops, a year before the existing *de facto* ban on such approvals expires.

Approvals stopped in October 1998, after public confidence in GM crops collapsed across Europe. Individual states can block new approvals, and the refusal of six nations — France, Italy, Greece, Denmark, Austria and Luxembourg — to cooperate in the granting of European Union (EU) licences brought the process to a halt. Only five GM crops — three varieties of maize (corn) and two of rape — are currently licensed by the EU, against a total of 53 approved worldwide.

The EU tried to end the impasse this February by adopting stricter and more transparent rules for the use of GM crops. These require biotech firms to carry out risk assessment and long-term surveillance of their crops' effects on human health and the environment (see *Nature* **409**, 967–968; 2001). Further amendments on labelling and long-term traceability were added in July.

But this directive will only come into force

in October 2002. In a bid to resume approvals before then, commission officials plan to ask EU member states to approve new licences on the basis of voluntary commitments from the applicants, similar in scope to the compulsory ones featured in the directive.

If the new guidelines are agreed, the approval procedures could be resumed by the end of the year, says David Byrne, head of the commission's Health and Consumer Protection Directorate.

Member states backing the moratorium have reacted cautiously. "If the EU can guarantee the safety standards listed in the new directive, an early start is conceivable," says Alexander Haslberger, head of biotechnology at Austria's social-security ministry, which regulates the country's GM crops.

But even the new directive is unlikely to end Europe's fierce debate over the acceptability of GM crops. Environmental groups are still calling for a complete ban. Meanwhile, some US firms argue that the rules on labelling and risk assessment amount to trade restrictions, and have called for them to be challenged under World Trade Organization rules.

AP

European court rejects protest over biotech patenting

Munich The European Court of Justice in Luxembourg has dismissed an appeal by the Netherlands against the European Union's controversial biopatenting directive.

The directive, which was approved in 1998, controls the patenting of biotechnological inventions such as gene sequences, and chemicals extracted from plants. It has been opposed by environmental groups, many of which object to patents on parts of living organisms.

The Netherlands took legal action a few months after the directive was passed, claiming that the rules violate human dignity and lack a legal basis. Italy and Norway have backed the Netherlands' appeal, and France, Belgium and Germany have attempted to revise the directive's wording.

But the European Court judges ruled last week that the regulation of patents involving human material is tight enough to protect human rights. The directive clearly states that no parts of the human body in any stage of development can be patented, they argue. The directive could still face a difficult future, however. Only 4 of the 15 European Union member states have adopted the regulations.



Information blackout: websites carrying plans of US nuclear power stations have been shut down.

Nuclear power plants go offline amid terrorist fears

Washington The US Nuclear Regulatory Commission has joined the growing list of organizations that have closed down websites as a precaution against terrorism. The commission has withdrawn information about all 103 operating US nuclear plants, including precise locations and data on safety systems. The site also offered access to technical documents, engineering circuits and photographs, all of which were previously deemed non-sensitive.

The move follows similar decisions by the Environmental Protection Agency and the Centers for Disease Control and Prevention (see *Nature* 413, 558; 2001).

Japanese 'spy' not guilty, say compatriots

Tokyo Japanese scientists have rallied behind Hiroaki Serizawa, one of two researchers accused of stealing DNA constructs and cell-line samples from a US research centre.

Serizawa and fellow researcher Takashi Okamoto are accused of taking the samples from the Cleveland Clinic Foundation in Ohio in 1998 (see *Nature* 411, 225–226; 2001), and of passing them to the Institute of Physical and Chemical Research in Tokyo, where Okamoto now works. Serizawa, now at the University of Kansas Medical Center, has been charged with espionage by the US Federal Bureau of Investigation.

Serizawa's supporters, including Kenichi Arai and other senior figures at the University of Tokyo, met this week in Tokyo to discuss how differences between Japanese and US research cultures could have caused the incident. The event doubled as a fundraiser for Serizawa's legal costs.

♦ www.innocentserizawa.org

Journal editors defect in protest at subscription costs

Washington Forty members of the editorial board of the *Machine Learning Journal* have resigned in a protest over subscription

charges. All have defected to the rival *Journal of Machine Learning Research*, the online version of which is available free of charge.

They argue that the subscriptions charged by Kluwer, publishers of the *Machine Learning Journal*, hinder communication within and outside their field. Institutional and personal subscriptions cost US\$1,050 and \$120, respectively.

The *Journal of Machine Learning Research* was launched last year with the support of the Scholarly Publishing and Academic Resources Coalition, an alliance of research institutions and libraries that aims to drive down the cost of academic publishing. A quarterly print version is sold by MIT Press, based in Cambridge, Massachusetts.

♦ www.arl.org/sparc

Battle plan to beat BSE goes up in smoke

Tokyo Japan's strategy for tackling bovine spongiform encephalopathy (BSE) has been hit by a refusal by incinerator operators to burn the animal-feed products that are thought to spread the disease.

The first case of BSE in Japanese cattle was confirmed last month (see *Nature* **413**, 240; 2001). The government has since ordered the destruction of all meat and bone meal products, the animal feed made from other

animals which was behind the spread of the disease in Britain.

The government aims to destroy some 200,000 tonnes of meat and bone meal, but incinerator operators have refused to cooperate, citing concerns about local residents' reactions and the safety of workers, who might inhale the waste. Meanwhile, the government has reportedly authorized the disposal of the meal by mixing it into cement, claiming that the high temperatures involved in cement production would destroy the protein that causes BSE.

US researchers benefit from data-crunching dynamo

San Diego A new supercomputer capable of carrying out five trillion operations per second is online at the Lawrence Berkeley National Laboratory in California.

The IBM RS/6000 SP system at the laboratory's National Energy Research Scientific Computing Center began working in late August. It is named Seaborg in honour of Glenn Seaborg, winner of the 1951 Nobel Prize in Chemistry and former director of the laboratory.

The centre is the US Department of Energy's biggest unclassified computing resource, and can be used by more than 2,000 scientists around the country. Seaborg is

already being used in climate-modelling and astrophysics studies.

♦ www.lbl.gov

Computer chat lines fail to seduce judges

London The prestigious Loebner Prize gold and silver medals — awarded to computers that convince judges that they are human during a five-minute conversation — have yet again gone unclaimed. The medals, worth US\$100,000 and \$25,000 respectively, have never been won in their 11-year history.

After chatting with eight computers and two humans at London's Science Museum,



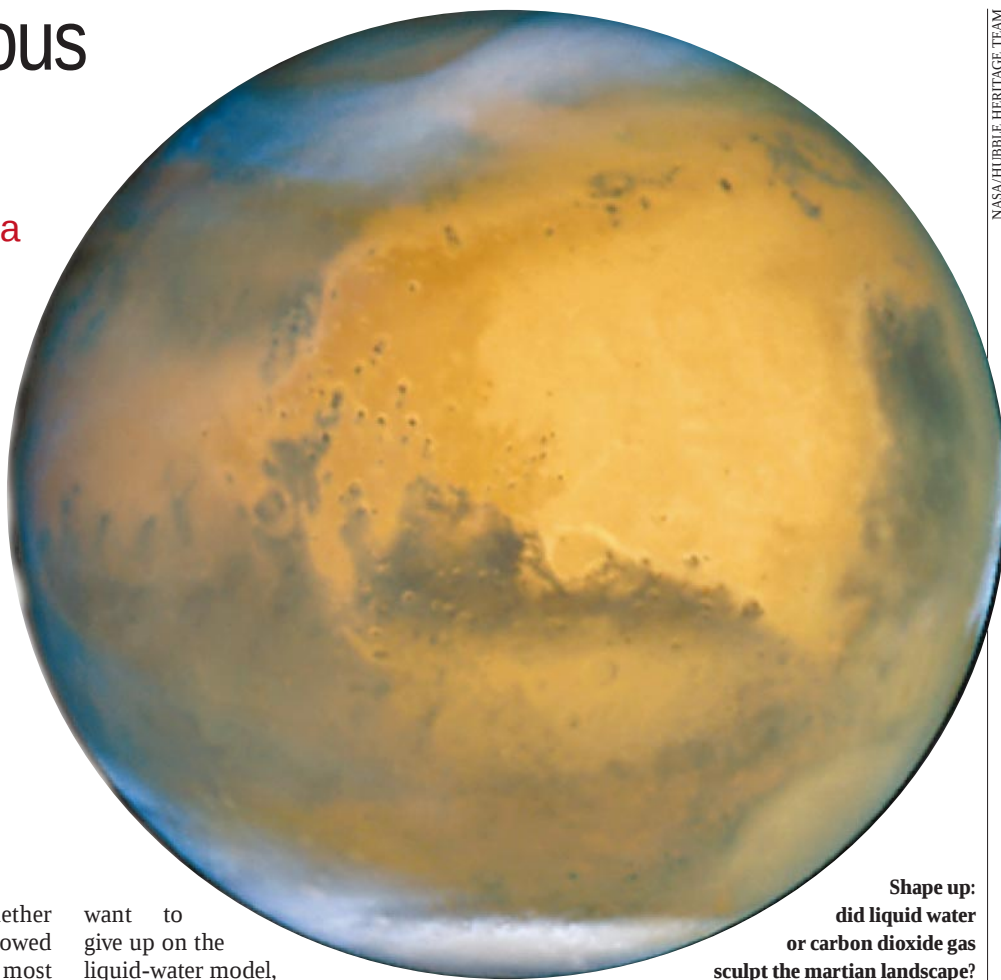
Bronze medallist Richard Wallace.

the judges awarded the \$2,000 bronze medal for the best program to ALICE, created by Richard Wallace, founder of the ALICE Artificial Intelligence Foundation in San Francisco. Despite answers such as "Leo said I be capable of learning therefore he classified I as an neural system", one judge ranked ALICE higher than one of the humans, says Wallace.

♦ www.alicebot.org

The outrageous hypothesis

For over a quarter of a century, planetary scientists have believed that water helped to shape the surface of Mars. Now one geophysicist is trying to prove them wrong. Larry O'Hanlon reports.



NASA/HUBBLE HERITAGE TEAM

When it comes to deciding whether or not liquid water once flowed over the surface of Mars, most geologists are happy to believe the evidence of their own eyes. The surface of the red planet is marked by channels and gullies that are virtually identical in form to water-created landscapes on Earth¹. As the saying goes, if it looks like a duck and quacks like a duck, it must be a duck.

But when Nick Hoffman, a geophysicist from La Trobe University in Victoria, Australia, discusses the arguments for water on Mars, he rounds off his presentations with an image of a duck-billed platypus. It is a ruse that draws chuckles from planetary scientists, but Hoffman is not joking. According to his hypothesis, water has had little or no importance in shaping the martian landscape. Instead, his 'White Mars' model contends that gurgling, frothing and sometimes explosive venting of carbon dioxide gas from beneath the surface created the features.

With 25 years of Mars research focusing on water, Hoffman is having to work hard to get his ideas taken seriously. His initial publication, in August last year², raised eyebrows but won him few converts. Since then, however, his theory has gained momentum. Researchers who saw him present his ideas on a recent tour of the offices of NASA and the US Geological Survey say they were impressed. And data from NASA's Mars Odyssey orbiter, which will arrive at the red planet next week, should add to the debate. Scientists might not

want to give up on the liquid-water model, but Hoffman's hypothesis seems to be emerging as a credible rival.

"I would say that Nick has generated an outrageous hypothesis," says Jeff Kargel of the US Geological Survey in Flagstaff, Arizona. "But I don't mean that in a negative way. The outrage is to an existing theory. Even if it's totally wrong, it makes you think about Mars in a different way."

According to conventional wisdom, Mars has witnessed at least two periods during

which liquid water existed on its surface³. Present-day Mars, where average temperatures are below freezing and the atmospheric pressure is less than 1% of that on Earth, is incapable of supporting liquid water. But frozen water is present on and below the surface, and the atmosphere contains a small amount of water vapour. If the martian atmosphere had been denser in the past, the greenhouse effect could have allowed liquid

Shape up:
did liquid water
or carbon dioxide gas
sculpt the martian landscape?



Don't duck the issue: Nick Hoffman is challenging his peers to abandon their assumptions about Mars.

water to flow across the surface and carve out the distinctive valleys and channels.

The first of these wet periods is thought to have occurred around four billion years ago. Valleys, similar to those created on Earth by rivers, criss-cross areas of Mars that are believed to date from this time.

At least one similar period is thought to have occurred since then, probably around one billion to three billion years ago. Mars' vast flood channels, which are orders of magnitude bigger than those on Earth, probably formed during this time. Researchers think that water flowed from equatorial regions towards the planet's northern hemisphere, although the source of the water and the details of the water cycle remain unclear.

What is clear is that Mars must have gained and lost an atmosphere on at least two occasions for these wet periods to have occurred. How this happened is the subject of speculation, but researchers have identified several potential mechanisms³. Heat-trapping gases, mainly carbon dioxide, are thought to have been released from frozen underground stores by surges of heat from inside the planet. But the resulting atmosphere may have been inherently unstable, with carbon dioxide being removed and trapped in rocks and ice. A change in Mars' magnetic field may also have played a role, exposing the planet to the stream of charged particles known as the solar wind and allowing it to strip away the atmosphere.

The images of gullies and channels, together with suggestions as to how the planet could have gained and lost an atmosphere, provide a good argument for the presence of liquid water on Mars. But not all of the data fit with the idea. NASA's Mars Global Surveyor, which is currently in orbit around the planet, has generated many new images, some of which have been touted as evidence of water on Mars. For every new piece of this evidence, however, there are complications.

Whither water?

Mars Global Surveyor has, for example, been unable to detect any carbonate minerals — a necessary product of liquid water interacting with Mars' abundant carbon dioxide². The craft's high-resolution images have also failed to reveal the expected finer features of water, such as the tributaries that would have fed the larger drainage features, and the shorelines of putative lakes and oceans⁴. And the flood channels lack obvious sources, as there is no evidence of large lakes or other reservoirs at the heads of the channels³. "There are a lot of problems with the water model," admits Michael Carr of the US Geological Survey in Menlo Park, California.

Problems like these leave room for debate, and Hoffman, for one, believes that the liquid-water model needs replacing. In his theory, Mars may always have had a cold, low-pressure atmosphere. Rather than being shaped by water, he claims that the distinctive features

were formed by eruptions of carbon dioxide.

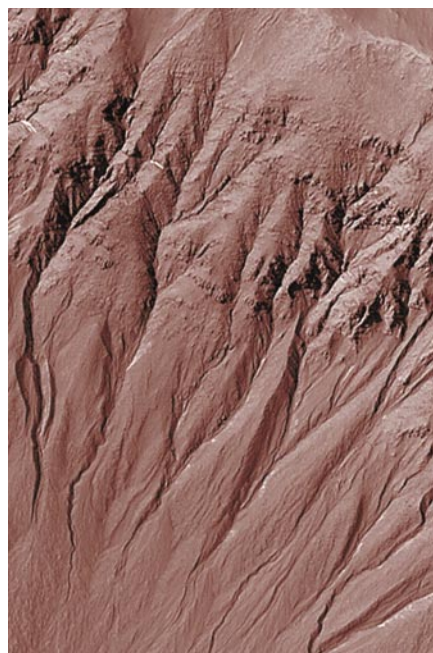
In the case of the flood channels, Hoffman proposes that the process began when steep slopes on the martian surface were heated from inside the planet. This melted underground deposits of frozen carbon dioxide, which is thought to make up around 5% of the upper layers of Mars. The melting weakened the slopes, which cracked under their own weight, exposing solid carbon dioxide within the rocks to the atmosphere.

Hoffman compares the creation of these cracks to the opening of a can of fizzy drink that has been violently shaken. Because the pressure of the martian atmosphere is so low, the carbon dioxide would have flashed violently into a gas, bursting from the rocks as the fizzy drink bursts from the can.

Together, the cracking of the slope and the explosion of carbon dioxide would have created an avalanche of gas and tumbling debris. Rocks within this turbulent flow would have ground against each other, exposing more carbon dioxide stores and prompting further explosions. The flow would eventually have ground to a halt when the stores of frozen carbon dioxide were exhausted, by which point, Hoffman argues, it could have travelled thousands of kilometres.

These frothing flows may sound odd, but Hoffman backs them up with established facts

For a long time, people have been viewing Mars through 'blue-tinted glasses'.



Martian mimic: in the 'White Mars' model, explosions of gas created gullies and channels that resemble those generated by water on Earth.

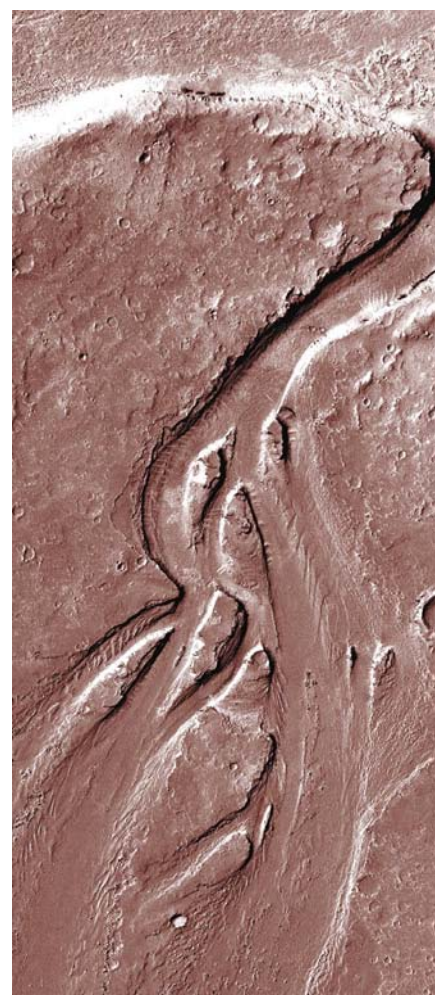
about the planet. Orbiting spacecraft have directly observed large amounts of solid and gaseous carbon dioxide on Mars³. Except for some ice at the poles and a tiny amount of water vapour in the atmosphere, the same cannot be said of water. And the melting and exploding of carbon dioxide can be explained in terms of the pressures and temperatures of the present-day martian atmosphere².

Closer to home

Such facts make Hoffman's ideas seem plausible, but they offer no evidence as to whether the flows actually occurred. It is this gap in the story that has generated the most scepticism. But doubt is starting to diminish now that Hoffman has presented his ideas on what he claims are analogous processes on Earth.

Volcanic eruptions provide one example. Hot ash and lava can travel tens to hundreds of kilometres on a cushion of gases released from volcanoes. Undersea avalanches can also trigger flows of mud, sand and boulders which, supported by the surrounding water, can travel thousands of kilometres.

These analogies seem to have helped Hoffman's cause. "He made a major forward stride when he gave a talk at the US Geological Survey and showed a submarine flow," says Kargel. "It's something that we can't dismiss."



NASA/JPL/MALIN SPACE SYSTEMS

Hoffman's ideas may also explain an aspect of the martian landscape that has stumped advocates of the liquid-water model — the absence of headwaters at the sources of the flood channels. The source areas look like blocky, jumbled landscapes where the underlying ground suddenly lost volume and crumbled. Other researchers have shown that pressurized carbon dioxide exploding out of the ground could have left behind just such a landscape⁵.

Out in the cold

Since his original publication, Hoffman has extended his model to include another martian feature that is commonly associated with water — the gullies seen in the walls of craters. Michael Malin and Kenneth Edgett of Malin Space Science Systems in San Diego, California, attracted substantial press interest last year when they argued that the gullies, which are found in high-latitude craters in the planet's southern hemisphere, look sharp-edged and newly cut — perhaps just years or tens of years old⁶. Malin and Edgett suggest that the gullies are the result of very infrequent conditions that allow frozen ground-water to seep from the tops of the craters.

But Hoffman says that this theory shows just how entrenched and irrational the liquid-water model has become. "We know from the atmosphere of Mars that there are not multiple sources pumping water vapour into the atmosphere all over the southern hemisphere," he says. He also points out that it is difficult to explain why the gullies are found only on the colder, poleward-facing slopes of craters instead of on the sunnier sides, where there would be more energy available to melt the underground water.

Hoffman's explanation, which is similar to one by Donald Musselwhite and colleagues at the University of Arizona in Tucson⁷, contends that the gullies were etched out by a miniature flow of gas and debris akin to that which created the flood channels. In his model, the process begins when sunlight heats

up the carbon dioxide snow. Just like water snow on Earth, this melts from the bottom upwards, because sunlight shines through it and warms the ground underneath. On Mars, the melting generates carbon dioxide gas, which builds up under the snow, eventually bursting out when the pressure becomes too great. "If you were there you'd see this thing go by like a little miniature avalanche," claims Hoffman. "It's a fluidized flurry of debris and gas boiling down the hillside into a channel."

He backs up his argument by showing a series of images of the same gully in different seasons. In winter, the whole scene is covered in carbon dioxide snow. Dark patches appear as spring nears and the snow sublimates away to reveal the ground. But at one critical point, the shaded gullies can be seen cutting across the thinning snow. That, says Hoffman, is the moment at which they are active. The gullies appear only on poleward-facing slopes, he argues, because more snow accumulates there. The snow on the southern slopes, in contrast, is too thin to generate avalanches.

Staking a claim

As Hoffman develops his arguments, planetary researchers are beginning to admit that they cannot simply dismiss them. "No one could come up with an argument that could really drive a stake into the heart of it," concedes Carr, who heard Hoffman discuss his ideas at a meeting in August at NASA's Lunar and Planetary Institute in Houston, Texas.

"If water wasn't involved and it was all created by carbon dioxide, that would be an amazing discovery," agrees Victor Baker of the US Geological Survey's Flagstaff office. But, he cautions, it would also entail an astounding mimicry of water-created features.

Hoffman suggests that some of the resistance stems from a desire to view Mars as an Earth-like place. In the early twentieth century, the idea that Mars was criss-crossed with canals made by intelligent beings was popular. Ever since then, says Hoffman, people have been viewing the planet through "blue-tinted glasses", writing about and studying a Mars defined by terrestrial imaginations, rather than the Mars it actually is.

But others in the field are less impressed with Hoffman's ideas, and say that their objections are based on science, not tradition. "It doesn't strike me as plausible," says Bruce Jakosky, a planetary scientist at the University of Colorado in Boulder.

He points to good evidence that Mars contains a substantial amount of water³ and that analysis of some martian meteorites⁸ shows that it may have had periods of warm and wet climate. Water vapour in the atmosphere contains more deuterium, a heavier isotope of hydrogen, than water found on the surface, as water molecules containing the heavier isotope are less likely to be lost to space. But analysis of some martian meteorites suggests that these higher levels were also present



Big picture: Bruce Jakosky argues that the liquid-water model provides a more coherent overview.

during some periods on the surface, indicating that atmospheric water vapour may have condensed and fallen as rain to the surface.

Perhaps Jakosky's most important criticism is that Hoffman is attempting to explain individual features one at a time, rather than incorporating his explanation into a larger model, as has been done with the liquid-water theory. "The history of the atmosphere, history of geology, history of the crust — that's the way to do it," says Jakosky.

At the moment, it seems that both models fit with the information available. Those data, however, are likely to be significantly enhanced in the coming year, as Mars Odyssey starts its work, looking for water-altered minerals and analysing the composition of volcanic rocks. All of this could say a lot about the role, or lack of one, that liquid water had in sculpting the planet's surface.

Hoffman says his theory would be bolstered if Mars Odyssey fails to find large amounts of carbonates. But not finding these minerals is unlikely to make researchers drop the liquid-water model. Like much of the data, analysis of carbonates provides only circumstantial evidence. And because definitive experiments, such as drilling deep into Mars in search of liquid water, are technologically and financially unfeasible, the arguments seem set to run and run.

In the meantime, Hoffman may have to be content with using his duck-billed platypus to persuade planetary scientists to re-evaluate their theories. Whether or not his outrageous hypothesis is correct, pushing researchers into this unfamiliar territory will benefit everyone's understanding of Mars. And that, for most, makes his ideas worth considering. ■

Larry O'Hanlon is a freelance writer in Cool, California.



What a gas: Victor Baker says it would be amazing if carbon dioxide created Mars' features.

It all falls into place...

Researchers working on molecular self-assembly have never lacked ambition, but their dreams of producing commercially viable devices always looked like a distant goal. That may be about to change, says Philip Ball.

Imagine a future in which computers build themselves. The nanoscale components of these machines would simply be brought together in solution and stirred gently. By tweaking the chemistry of the components so that some are attracted to each other, while others are repelled, the individual parts would assemble themselves, as if by magic, into a working whole.

Such is the hype generated by nanotechnology enthusiasts. But researchers working in this field have always regarded self-assembling computers as a distant dream. The chances of devising a commercially viable means for creating sophisticated electronic circuitry that builds itself remain remote.

But as the field evolves, more modest practical applications for self-assembly are emerging sooner than experts had dared hope. Some researchers are focusing on how to combine self-assembly with existing manufacturing methods. Others are developing new gene-screening techniques by using DNA to guide the self-assembly process. There has been little fanfare, but a smattering of start-up companies indicates that the ideas have commercial potential. "Self-assembly is going to be a big deal," predicts George Whitesides, a Harvard University chemist who has done some of the field's pioneering work.

Causing a stir

Much of the inspiration behind self-assembling systems comes from biology, where intricate molecular machines put themselves together without any external control. The ribosome, the site for protein synthesis in the cell, is a classic example. Human ribosomes are made up of about 80 separate proteins and four strands of ribonucleic acid. In common with many natural and artificial self-assembling



systems, these components are held together by attractive forces which, compared with the bonds between the atoms within the components, are relatively easy to break.

Certain substances, such as detergents, can disrupt these forces, causing the ribosome to fall apart. But remove the detergent and the parts reassemble in the correct configuration, leaving a fully functioning ribosome. It is as if a dismantled watch could be rebuilt simply by throwing the parts together in a beaker of water and stirring.

The structures of the different components of the ribosome are key to this process. Each part is made up of a series of different groups of chemicals, and these interact with groups on other components in a well-defined way. The attractions and repulsions between these groups are such that the arrangement for a properly functioning ribosome is also the most stable one. Once recombined, the components naturally fall into this state.

The challenge for chemists working on self-assembly is to use their knowledge of molecular attraction and repulsion to devise artificial systems that will behave in the same way. And although mimicking the full glory of the ribosome remains a long-term goal, researchers have already created structures similar to simpler, repetitive biomolecular assemblies — such as the arrays of proteins that make up microtubules, part of the cellular skeleton, or the sheets of lipid molecules that line up to form cell membranes.

Whitesides and his Harvard colleagues are behind some of the



Stuck up: tiny gold tiles can be made to self-assemble in water into a range of structures by coating either the edges (above), faces (below) or both (left) with a water-repellent compound.

G. WHITESIDES

most distinctive work. They have created a range of repeating structures from simple units by coating particular surfaces of the units with substances that are either hydrophilic or hydrophobic (attracted to or repelled by water). When combined in water, the hydrophobic surfaces of the components tend to come together so as to minimize their contact with the surrounding liquid.

This July, Whitesides described how his team had applied this approach to flat hexagonal plates of gold just 50 nanometres thick and 10 micrometres wide¹. Some parts of the plates were coated with chromium, which is hydrophilic. Others were coated with a film of hydrophobic molecules. The Harvard team mixed hundreds of thousands of the plates in water, and allowed them to self-assemble over a few days.

The structures that formed depended on where the coatings were applied. When the edges of the hexagons were made hydrophobic, the plates assembled into a sheet of hexagonal tiles. But if the hexagonal faces were made hydrophobic, the plates formed columns, like stacks of coins. If both faces and edges were coated, the result was a cluster of columns sticking side by side.

The hydrophobic forces are strong enough to hold the assemblies together, but not so strong that initial misalignments or partial overlaps of surfaces get fixed in place.

Most of these irregularities are eventually corrected by the constant motion of the surrounding water and the plates themselves. By coating the hydrophobic surfaces with adhesives that set on exposure to either heat or light, the Harvard team managed to make the structures permanent.

The group has used a related technique to assemble log-shaped plastic components into miniature towers. Selected regions



George Whitesides has created 3D electronic circuits.

of the millimetre-sized logs were first coated with a bismuth alloy. The logs were then placed in water at a temperature above the alloy's melting point — 47 °C — and gently stirred. As the logs came into contact with each other, areas of molten alloy stuck together. By applying the alloy to specific parts of the plastic components, Whitesides' team produced a system that self-assembled into a tower of criss-crossing logs².

Whitesides has also managed to create working electronic circuits out of millimetre-scale plastic blocks. Some sides of the blocks had components, such as wires or light-emitting diodes, printed onto them. Others were designed so that they would join up with neighbouring blocks in specific ways. When gently agitated in water, the blocks spontaneously assembled into a working three-dimensional electronic circuit³.

Whitesides' circuits are still a long way from being a self-assembling computer, and the components needed a lot of careful hand-tooling, contrary to the ultimate aims of self-assembly. But his team has proved that working circuits can be self-assembled. If the researchers can repeat their work using blocks of micrometre dimensions, which are difficult to manipulate manually or robotically, they could point to way to building a new breed of ultra-small circuits.

Building blocks

In the short term, rather than using such techniques to create a completely new kind of manufacturing process, most researchers believe that self-assembly methods will be combined with standard chip-building technologies. This way, they argue, the convenience and low cost of self-assembly can be exploited where appropriate, leaving established methods to do the rest.

Last year, for example, Bob Wolkow and his colleagues at the Steacie Institute for Molecular Sciences in Ottawa, Canada, showed that styrene molecules will self-assemble into rows on a silicon surface that has been coated with hydrogen atoms⁴. The

The emergence of new applications has happened sooner than expected.

process is triggered by using a scanning tunnelling microscope to remove a single hydrogen atom from the surface, which creates a site at which a styrene molecule can bind. The binding process removes an adjacent hydrogen atom, setting off a chain reaction in which styrene molecules are added in a line.

Wolkow's work is some way from a commercial application, but if similar rows could be built from molecules that can transfer electrons, they could form the basis for wires, just one molecule wide, that would self-assemble on the surface of a pre-manufactured silicon chip.

Attractive forces

Others predict that standard manufacturing methods could be used to draw a circuit blueprint on the surface of a chip, and self-assembly techniques used to ferry prefabricated components to the appropriate places in that circuit. Angela Belcher of the University of Texas at Austin has recently identified a series of peptides — fragments of protein molecules — that attach themselves selectively to materials used in electronics, such as gallium arsenide⁵. Together with



Angela Belcher hopes to self-assemble chips.

researchers at the University of California, Santa Barbara, she is in the process of setting up a company called Semzyme to exploit the idea, and is now searching for double-headed peptides that could bind circuit components to the surface of a chip.

Chris Murray, head of the nanoscale materials and devices division at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, thinks that the integration of self-assembly with conventional manufacturing techniques could have commercial applications. "We hope that self-assembly will be able to inexpensively replace certain stages in the production of materials and devices, where control is needed at the molecular level," he says.

The new commercial initiatives extend beyond microelectronics. Chad Mirkin and colleagues at Northwestern University in Evanston, Illinois, are exploiting the way in



Gene screen: Chad Mirkin has launched a company to use self-assembly in genetic tests.

which DNA binds selectively to complementary sequences in the hope of developing a new technique for genetic screening⁶.

Mirkin first identifies the sequences of bases at either end of the DNA he hopes to screen for. He then creates two sets of strands of DNA, each complementary to one of the 'end' sequences of the target DNA. These are attached to gold nanoparticles, with each nanoparticle linked to just one type of strand, and the nanoparticles are dispersed in water.

If the target DNA is added to the solution, gold nanoparticles bind to both ends of the sequence using the complementary strands. The gold particles cluster together, eventually forming a large enough group to scatter light and change the colour of the solution. Last year Mirkin launched a company called Nanosphere, based in Northbrook, Illinois, to commercialize the idea.

Such emerging companies are evidence of the field's gradual move towards commercial applications. At the same time, the more blue-sky projects undertaken by Whitesides show that self-assembly has not lost its visionary zeal. Many in the field are still intent on creating the nanoscale chemical sensors or mechanical devices that are difficult or impossible to make by other methods. But success for the start-up companies should provide inspiration to all quarters of self-assembly research.

Fraser Stoddart, who works on self-assembling electronic devices at the University of California, Los Angeles, says that the field is moving a lot faster than he expected. "Something that I thought would only ever be a dream in my lifetime now stands a good chance of becoming a technological reality before this decade is out," he says. ■

Phillip Ball is a consultant editor of Nature.

1. Clark, T. D., Tien, J., Duffy, D. C., Paul, K. E. & Whitesides, G. M. *J. Am. Chem. Soc.* **123**, 7677–7682 (2001).
2. Oliver, S. R. J., Clark, T. D., Bowden, N. & Whitesides, G. M. *J. Am. Chem. Soc.* **123**, 8119–8120 (2001).
3. Gracias, D. H., Tien, J., Breen, T. L., Hsu, C. & Whitesides, G. M. *Science* **289**, 1170–1172 (2000).
4. Lopinski, G. P., Wayner, D. D. M. & Wolkow, R. A. *Nature* **406**, 48–51 (2000).
5. Whaley, S. R., English, D. S., Hu, E. L., Barbara, P. F. & Belcher, A. M. *Nature* **405**, 665–668 (2000).
6. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. *Nature* **382**, 607–609 (1996).

Owen was right, as Darwin's work continues

Tribute to Darwin was not a veiled attack but a genuine expression of hope for the future.

Sir—In his Words essay “Owen’s Parthian shot” (*Nature* **412**, 123; 2001), Kevin Padian discusses a letter from Richard Owen to Spencer Walpole, a politically important member of the British Museum’s Board of Trustees, shortly after Charles Darwin’s death in 1882. Owen had been asked whether Darwin’s scientific accomplishments merited a statue in the new museum. Owen, noting that the decision should be left to the museum’s administrators, compared Darwin’s singular contributions to biology to those of Copernicus to astronomy. He predicted that Darwin, like Copernicus, would have successors who would complete a fuller understanding of the concepts he had initiated.

Padian claims that in this post-mortem analysis and comparison, Owen was aiming “a Parthian shot at his rival in the guise of a tribute—a masterful and delicate duplicity”. Duplicity? Hardly. It was an analogy Owen frequently used. The letter reproduced in full in Padian’s

essay was in fact the last, as far as we know, in a series of four similar letters that began in 1877. In each of these, Owen uses the same analogy to restate figuratively the basic objections he had expressed when Darwin’s *Origin of Species* was first published in 1859: that although natural selection is a valid mechanism to explain species diversification through time, it did not answer the more basic question of the origin of the inheritable individual differences subsequently “naturally selected” for survival in a surrounding and changing environment.

Without an answer to the problem of inherited variations, Owen believed that the origins of species were not fully understood. Darwin himself confessed: “Our ignorance of the laws of variation is profound.” With his copernican analogy, Owen was expressing his confidence that in the coming years, as with the heliocentric theory, Darwin’s theory of evolution would find its own Galileos,

Keplers and, finally, its Newton.

The twentieth century belonged to the geneticists, who shifted the field from the species level to that of the individual. Ironically, the 1908 meeting in Cambridge to celebrate darwinism provided a platform for the practitioners of the newly named science of genetics, led by William Bateson—those “snarling dogs”, complained one loyal darwinian—“to lower Darwin and elevate themselves”.

This meeting marked the beginning of the dominance of genetics, with its mutations and its mathematics, leading towards the productive evolutionary syntheses 30 years later.

Among that vast body of creative workers Darwin’s work had spawned there were, and are, as Owen’s analogy predicted, his Keplers, Galileos and perhaps a Newton.

Jacob W. Gruber

Department of Anthropology, Temple University,
Philadelphia, Pennsylvania 19122, USA

SPOrt to fill a gap on cosmic maps

Sir—Your well-written and pertinent News Feature (*Nature* **411**, 880–881; 2001) “Raking through the embers” omits to mention the Sky Polarization Observatory (SPOrt), which is a space mission currently under development that is precisely aimed at a sky-polarization survey.

The programme was proposed to the European Space Agency and was selected in 1997, and is now fully funded by the Italian Space Agency (see <http://sport.tesre.bo.cnr.it>). SPOrt will be on board the International Space Station for at least 18 months from 2005. It is designed to directly measure Stokes’ *Q* and *U* parameters at the microkelvin level, providing an accurate measurement of the polarization of the cosmic microwave background (CMB) on a large angular scale ($> 7^\circ$) from four independent channels in the frequency range 22–90 GHz. More than 80% of the sky will be scanned under the most favourable conditions for testing re-ionization models.

Complementary to the MAP and Planck projects discussed in your feature, which will map the CMB anisotropy with unprecedented accuracy, SPOrt should also provide the much-desired galactic foreground polarization maps so far unobserved at frequencies above a few

gigahertz, and study the correlation between the available COBE and MAP temperature maps and polarization patterns over large sky regions.

Stefano Cortiglioni

Principal investigator, SPOrt project,
Istituto di Tecnologie e Studio delle Radiazioni
Extraterrestri, Consiglio Nazionale delle Ricerche,
Via P. Gobetti 101, 40129 Bologna, Italy

Singapore considers institute’s future

Sir—Your News Feature “Building a biopolis” (*Nature* **412**, 370–371; 2001) quotes Philip Yeo, chairman of the Singapore National Science and Technology Board, as saying that the Institute of Molecular Agrobiolgy (IMA) has been “a criminal waste of taxpayers’ money”.

While respecting Yeo’s personal view on resource allocation, the Ministry of Trade and Industry, which oversees the board, does not agree with his view. The IMA has done good work and has received international recognition.

Singapore has decided to make the biomedical sciences the focus of its life-science research efforts. The Ministerial Committee on the Life Sciences is now conducting a review of how the IMA should be repositioned with

this in mind, and will make a decision later this year, after consulting its international advisers.

Khaw Boon Wan

Ministry of Trade and Industry, 1000 High Street,
09-01 The Treasury, 179434 Singapore

Peers under review

Sir—Your News Feature “Peers under pressure” (*Nature* **413**, 102–104; 2001) makes me glad that I work in “geology ... [where] complaints are few and far between”. I sign all of my reviews, positive and negative, and people still talk to me.

Some of the cited examples of foul play by reviewers have obviously had major implications for tenure, patent rights and so on. No wonder, then, that some are tempted to seek a legal solution to some of the injustices that the peer-review system has enabled the less scrupulous to perform.

Problems could doubtless arise in the process of separating genuine complaints from the paranoid or vindictive, and the accused might launch countersuits and writs. But a few successful prosecutions might well encourage wrongdoers to consider changing their ways.

Stephen K. Donovan

Department of Palaeontology, Nationaal
Natuurhistorisch Museum/Naturalis, Postbus
9517, 2300 RA Leiden, The Netherlands

The best and worst of times

What winners will emerge from the battles over access to scholarly data?

David R. Worlock

The Internet has caused a revolution in research. Scientists now work remotely from one another but in tightly organized groups, with an immediacy of communication that demands instant availability of the corpus of knowledge on which their work is based. In the process, the underpinnings of a system of filtering and selecting results that has been effective for the past 150 years have been questioned and shaken, and the roles of the actors — publishers, librarians, scholarly societies and universities, as well as researchers — are being redefined.

This is not a picture of the manoeuvrings of powerful vested interests confined to a cul-de-sac in the progress of communication, as some have suggested. Rather, it is the proving ground for the future as we evolve from an information-rich to a knowledge-rich society. The patterns of behaviour created by scholarly research publishing will continue to evolve, setting the standards for communication in the business and professional worlds.

Let battle commence

When 28,000 researchers this year signed the Public Library of Science petition (www.publiclibraryofscience.org) demanding that publishers post scholarly articles on freely available, centralized servers six months after publication, a battle — both real and virtual — had begun. The clamour has been strongest in the 'hard' sciences. Andrew Odlyzko, in the current *Nature* web debate (www.nature.com/nature/debates/e-access/Articles/odlyzko.html), reports a story: following an e-print article about a new high-temperature superconductor, "every superconductivity laboratory in the world immediately began to make measurements on this new material and dash into print. Fifty e-prints had been posted on the web by the end of February — before the original paper was even published." In other words, conventional publishing is too slow. There is an inexorable pressure to connect the e-print servers of the world and to create a two-speed scholarly communication economy.

Do these changes undermine traditional, paper-based journals? Scholars still rely heavily on peer-reviewed publication. As authors, of course, they insist on appearing in the most prestigious, branded outlets, even if, as users, their views can be subtly different. Major studies in the United States (see overleaf) indicate that, although print is not under threat, a culture of reading outside core publications is growing; document-

delivery ordering is increasing by as much as 7.7% per annum (Association of Research Libraries research; www.arl.org). This startling growth is occurring as librarians under budgetary constraints reduce subscriptions and use interlibrary loan facilities to the fullest. Indeed, scholars may now be obtaining as many articles from ordering systems as they used to read via their institutions' subscribed publications.

The university and the library are being sharply redefined, in both the terms and style of access by scholars, and in the ownership of tradeable intellectual property in the form of copyrights. Some universities — the Massachusetts Institute of Technology is in the lead — already sell courses on the web. The idea of university consortia trading in the intellectual property created by their researchers is not far behind. And it follows that, if distributed storage but centralized searching is the model, libraries have no role.

Although subscriptions to printed journals still form 35% of average library budgets, document acquisition is 8% and rising. Librarians are evolving into powerful network-resource administrators. Years of library consortia and licensing schemes underline the power of collective bargaining (some publishers report up to 60% of sales to consortia rather than individual library purchasers), and remind libraries that if they have the intellectual property they can be traders, too.

As in every period of rapid change, there are also losers. Outside the consortia, and in the less-developed world, a genuine poverty of access is emerging as never before, with the scholarly rich and poor divided sharply on access and on the ability to stay abreast of the fast-moving research base.

The world of print journals is underpinned by commercial publishing, either via branded journals or by contract publishing for scholarly societies, whose income derives from journal subscriptions. Surprisingly, it is hard to determine the total number of journals: it could exceed 50,000, but there is a profitable core of 12,000–15,000, with publishers maintaining their profit margins by increasing prices when subscription numbers fall.

Rapid consolidation creates economies of scale, prevents players from being cut out by new intermediaries, and accounts for the reinvestment forced on publishers by expen-

sive digital conversion and new formats such as XML. The big are likely to get bigger and the small squeezed.

Publishers' response

How are publishers coping with the changing behaviour and demands of researchers? After a period of denial, a collaboration emerged in the late 1990s that would have been unthinkable a generation earlier. The fruits were the Digital Object Identifier Foundation (www.doi.org) and the CrossRef service (www.crossref.org) — attempts to use meta-data (data about electronic data) to improve access to journal articles wherever they were held in distributed publishing systems.

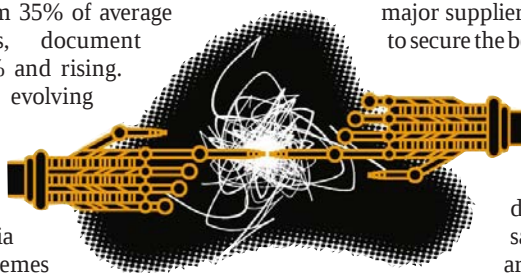
All this begs the ultimate question: who owns the customer? All publishers want to know the answer, so they can match changing habits. Smaller publishers may be able to use CrossRef's search facility to stay on researchers' radar, but librarians and administrators want to make smart deals with major suppliers and intermediaries to secure the best terms.

It may be too late. The Open Archives Initiative promises a lowest-common-denominator universal metadata standard (www.nature.com/nature/debates/e-access/

Articles/harnad.html). The Office of Scientific and Technical Information of the US Department of Energy is typical of 'big government' in desiring to provide unifying access standards across thousands of distributed e-print sites in universities. Scholars are looking to the next major change: knowledge representation and the addition of RDF (resource-definition framework) standards to current XML metadata. This will enable ontologies to be used that will allow searching on knowledge structures, not simply on terms and words, in turn creating a new standards debate, as domain-driven ontologies compete at the borders of their disciplines.

Meanwhile, the shape of the article itself as a reporting mechanism is changing. The addition of research files, results databases, software environments for running comparative results, laboratory videos and other materials heralds the day when the article becomes the core knowledge document to which archival and grey literature can be referenced and linked. ■

David R. Worlock is chairman of Electronic Publishing Services Ltd, 26 Rosebery Avenue, London EC1R 4SX, UK.



ALL IMAGES BY DAVID NEWTON

Lessons for the future of journals

Science journals can continue to thrive because they provide major benefits.

Carol Tenopir and Donald W. King

Technology is transforming all aspects of the journal system. Despite the current atmosphere of turmoil, more publishers are producing electronic journals, and the proportion of reading from electronic versions is increasing rapidly.

We have tracked the characteristics of US scholarly science journals from 1960 to 2000, received more than 15,000 scientist readership survey responses between 1977 and 2001, and collected cost data on the activities of authors, readers, publishers, libraries and other participants. By observing long-term trends and past successes and failures, we can learn much about the future of journals.

Our results show that, over these years, the journal system has remained surprisingly stable. Scientists still depend on scholarly journals for reporting research results, obtaining information and as reference sources. Also, the number of articles published per scientist, the amount of reading and the indicators of usefulness and value are virtually unchanged. The overall cost of the journal system has stayed relatively constant, although costs have shifted among participants.

The mistaken belief persists that new technologies can easily remedy any weakness in the system. All-or-nothing approaches — an entirely electronic system, or e-print servers or article databases totally replacing journals — have been advocated. This year, in particular, has seen bitter disputes (see below). But such disputes are not new and, if resolved thoughtfully, can benefit journals in the long term.

The journal system is at a critical stage — poor judgement could mean its deterioration or destruction. We believe that insights from the past can provide constructive guidance for innovators and the scientific community alike.

Communication technologies

In the 1960s and 1970s, many scientists believed that journal articles were written largely to satisfy tenure requirements and that publications were overpriced, too slow, unread, proliferating explosively and a huge waste of paper. Meanwhile, hundreds of studies on the channels of scientific communication found that authors write for many reasons, that growth in the literature is strongly correlated with the number of scientists, and that articles are generally well read (depending on the size of the audience) and considered extremely valuable inside and outside the academic community.

Many thought that electronic journals would resolve the flaws in the system. All the main components for an electronic-journal system existed by 1970 — authors could input digital text through magnetic tapes or cards; publishers used computerized photo-composition or computer-driven typesetting; libraries were beginning to automate; bibliographical databases were being searched online; scientists had computers and modems; and the foundations of the Internet had been established.

Yet a 1978 analysis by King Research forecast that it would be 20 years before electronic journals became commonplace, because acceptance of the necessary electronic processes would develop unevenly and slowly, as would their capabilities^{1,2}. In particular, standard codes for text input and better means of processing mathematical equations, chemical compounds and graphics were needed. The analysis predicted that the potential of electronic publishing would not be realized until all the components were fully in place and standardized.

These components continued to develop in the 1980s, and by the early 1990s, electronic journals finally emerged. Today, almost two-thirds of scientific journals are available both electronically and in print, and there are more than 1,000 electronic-only peer-reviewed journals (see www.arl.org/scomm/edir).

In addition to observing long-term trends of information-seeking and reading, we recently resurveyed scientists at two institutions to observe changes before and after electronic journals (Oak Ridge National Laboratory, 1984 and 2000; University of Tennessee at Knoxville, 1993 and 2001). We found that about a third of readings there are now from electronic sources; the amount read by these scientists is up (although time spent reading is about the same) and from a broader range of journals; scientists rely more on online search tools to find articles; the proportion of reading from personal and institutional subscriptions is unchanged; and there is more reading from copies of individual articles rather than whole issues of journals. Scientists continue to value highly the information they find in journal articles, whether print or electronic³.

Continuing innovation

High journal prices, continued perceptions of inadequacies in the journal system and fascination with new technologies have spurred a rash of innovative ideas for enhancing or replacing traditional journals.

Objectives of scholarly journals

- The journal system should serve as a means of communication of new, peer-reviewed and edited information. Thus, the information should be trustworthy and, to the degree possible, supported by other research findings.
- Journals and articles should be readily available to readers and accessible to an unlimited audience beyond the author's primary or immediate community.
- Because many articles are read years after publication, the system should provide permanent, locatable and retrievable archives for the information.
- The system should continue to provide alternative distribution means and media so that authors and readers can choose from alternatives that satisfy their specific needs, particularly to minimize time and effort.
- The system should protect against plagiarism, copyright-ownership violation and unauthorized modification or altering of the record of ideas, discoveries and hypotheses tested.
- The journals should properly convey the concept of prestige and recognition for authors, their research and their institutions.

From ref. 11

Examples are the Open Archives Initiative (www.openarchives.org), the Public Library of Science (www.publiclibraryofscience.org) and the Association of Research Libraries' SPARC initiative (www.arl.org/sparc). Many of these ideas show promise for improving scientific communication. Multiple options for distributing research results have strengthened scientific communication, and will surely continue to do so. But experience shows that it is much more difficult to adopt new technologies throughout a complex system than one might think.

The quality filters of peer-reviewing, editing, indexing, abstracting and bundling articles together in categories all add to the system's overall value. If we discard the functions performed by librarians and publishers, we will surely need to reinvent them. There is a tendency to promote an innovation as the only solution to a problem when, in fact, it would merely provide a niche



satisfying certain information needs.

History shows that a diversity of channels for distribution and publication increases the value of scholarly information. Publishers should not object to web archives, and authors should not abandon journals. Researchers should use multiple distribution channels, including self-archiving and publishing in traditional journals. Journals provide a stable archive of the literature, quality filters and other valuable aspects; web e-print servers allow quick access to more sources of information. Together, they serve the need of today's scientists for more knowledge from a wider variety of sources.

Publishers versus librarians

The 1960s and 1970s also saw efforts to develop central databases or archives of scientific articles, including preprints in physics and medicine. Preprint services were set up, notably the National Institutes of Health Information Exchange Group^{4,5}. But, despite extensive use, this failed through lack of funding. Even so, the growth in the distribution of individual articles was phenomenal — around 40 million copies in the United States by the late 1970s and more than 100 million today.

Libraries avoided the cost of buying infrequently read journals by obtaining copies of articles through interlibrary lending. The high cost of labour-intensive bibliographic verification and location, coupled with frequent unfilled requests, led to a proposal for a 'one-stop' US national periodical centre. But this was defeated in Congress, largely due to heavy lobbying by publishers and, oddly, some libraries. Its reported purpose was to replace the thousands of sources for obtaining copies of articles, although privately many felt that a central database would ultimately become electronic and replace the need for individual journals.

Publishers were concerned about the central database and convinced that they were being 'ripped off' by inter-library lending. A bitter and impassioned battle with librarians ensued, which was only partially resolved by the 1976 US copyright revisions. However, statistical studies^{6,7} subsequently revealed that interlibrary lending was not seriously affecting publishers' revenue. In fact, after the US Copyright Clearance Center (CCC) was established, revenue from royalty payments for lending was so low that the copyright requirements were broadened to cover all photocopying in an organization, which at least provided enough income to cover the cost of CCC operations. Factual information acquired earlier might have avoided much of this acrimony and expense.

The 1980s increase in journal prices has reopened the rift between librarians and publishers. Commercial publishers were called "the devil" at the Association of

College and Research Libraries' annual conference this year, and the Association of American Publishers complained of "the serious issue" of libraries giving away electronic copyrighted materials (*Washington Post*, 2 July 2001). The current *Nature* web debate (see www.nature.com/nature/debates/e-access) explores many conflicting viewpoints on the Public Library of Science's call for an author boycott of publishers, which started last month⁸. (It is too early to measure the effects of this action.)

Journal prices have certainly increased faster than inflation^{2,9}, whereas, over the same period, the publication cost per page for print journals, the overall publication costs per scientist and the costs per reading have decreased. This paradox is partly explained by the inherent nature of publishing costs. Publishers provide three basic article-related services: (1) collecting manuscripts of interest to readers in a discipline; (2) ensuring trust and quality; and (3) providing distribution and/or access. Many journals provide other useful information, such as editorials, letters, commentaries, reviews, guides to literature and database links.

Most publishers are reluctant to reveal their real costs, but we have derived a comprehensive model⁹. Essentially, journal publishing has large fixed costs, regardless of circulation, involving (1) and (2) above, other information-processing costs and overheads. To recover these costs, a low-circulation journal must charge more than a high-circulation journal. For example, a journal with 500 subscribers and typical costs must charge at least \$800 to recover fixed costs, but the same-size journal with 5,000 subscribers need charge only about \$110. It had been thought that electronic publishing would reduce costs (and prices), but print reproduction and distribution costs are typically only about \$25–35 per annual subscription. Electronic access avoids these costs, but has a substantial additional fixed cost — putting up full text on the web, staffing, software and other technology issues including design, functionality, searchability and speed. All of this significantly affects price.

Also, the numbers of articles and pages in journals have increased appreciably over the past 20 years, increasing the total costs of information content and distribution, despite reductions in the cost per page. Inflation and increased journal sizes together account for more than half of price increases. The combination of electronic and print versions has increased fixed costs and, therefore, prices. The average number of personal

subscriptions per scientist has roughly halved over 20 years, so publishers have had to recover their fixed costs by accelerated price increases to libraries, where demand is less price-sensitive (see ref. 9).

The fall in personal subscriptions does not mean that scientists are reading less, rather that they are using more library-provided articles, document-delivery services, interlibrary lending and full-text databases. An average university scientist reads three times more library-provided articles than in 1977, and scientists in other organizations average seven times more.

Despite increases in price and amount of reading, the total subscription revenue to publishers has remained relatively constant when normalized by numbers of scientists and amount of reading. University libraries have borne the brunt of the price increases because of the fall in personal subscriptions.

The extraordinarily high prices of commercial journals are due in part to their low circulation. For example, in 1995 the median circulation for commercial publishers was 1,400 subscriptions, compared with 5,600 for society publishers. Of course, commercial publishers tend to serve new, small scientific specialties. It is certainly not true of all journals or all publishers that high prices are due to exorbitant profits. Commercial prices are correlated with the number of titles published¹⁰. This suggests that market power achieved through size permits higher prices and profits, but larger, labour-intensive publishers almost certainly have higher overheads on direct costs. Yet being large means that publishers can do research and their own technical design. This is why projects such as BioOne (www.bioone.org) seek to bring these advantages of size to small publishers.

Benefits of knowledge

Publishers, scientists and librarians need to base their actions on facts, not emotions. Publishers are not a barrier to access. Much of the current ill will stems from ignorance — librarians and scientists do not understand the causes of rising journal prices, and publishers, just as with interlibrary lending, are afraid of becoming obsolete if readers have access to articles without paying for them. As before, there is no solid evidence that this will happen. One potential bridge between publishers and librarians is the expanding use of site licences, which can minimize the effects of factors that increase prices, and allow librarians and publishers to negotiate their differences. ■

Carol Tenopir is at the School of Information Sciences, University of Tennessee, Knoxville, Tennessee 37996, and Donald W. King is at the School of Information Sciences, University of



Pittsburgh, Pennsylvania 15260, USA.

1. King, D. W. & Roderer, N. K. *Systems Analysis of Scientific and Technical Communication in the United States: The Electronic Alternative to Communication Through Paper-based Journals* (King Research, Washington DC, 1978). Available from NTIS PB 281-847.
2. King, D. W., McDonald, D. D. & Roderer, N. K. *Scientific Journals in the United States: Their Production, Use, and Economics* (Hutchinson Ross Publishing, Stroudsburg, Pennsylvania, 1981). dwking@mail.sis.pitt.edu
3. Tenopir, C. et al. Scientists' use of journals: Differences (and similarities) between print and electronic. In *National Online 2001: Proceedings 469-482* (Information Today, Inc., Medford, New Jersey, 2001).
4. Aitchison, J. *Alternatives to the Scientific Periodical: A Review of Methods Reported in the Literature* (Office for Scientific and Technical Information, London, 1974).
5. Hills, J. A. *A Review of the Literature on Primary Communications in Science and Technology* (Aslib, London, 1972).
6. King, D. W. et al. *Library Photocopying in the United States* (US Government Printing Office, Washington DC, 1977).
7. McDonald, D. D. & Bush, C. G. *Libraries, Publishers and Photocopying*. King Research Report (US Copyright Office, Washington DC, 1982). Available from ERIC ED 226-732.
8. *Nature* **413**, 1-3 (2001).
9. Tenopir, C. & King, D. W. *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers* (Special Libraries Assoc., Washington DC, 2000). Available from <http://www.sla.org>
10. McCabe, M. J. *The Impact of Publisher Mergers on Journal Prices: A Preliminary Report* (Association of Research Libraries, Washington DC, 1998). Available from <http://www.arl.org/newsltr/200/mccabe.html>
11. King, D. W. & Tenopir, C. Scholarly journal and digital database pricing: threat or opportunity? In *Bits and Bucks: Economics and Usage of Digital Collections* (eds MacKie-Mason, J. & Lougee, W.) (MIT Press, Boston, in the press).

Travelling the genomic road

Pharmacogenomics

commissioning editor Katie Prickett
Ashley Publications. 4/yr (to be 6/yr from 2002). £425, \$745 (corporate); £210, \$365 (academic)

Physiological Genomics

editor-in-chief Victor Dzau
American Physiological Society. Variable frequency. \$200 (institutional), \$133 (individual), online access free this year

Current Genomics

editor-in-chief Stefan M. Pulst
Bentham. 4/yr. \$500 (print or online, academic); \$550 (academic, print & online); \$88 (individual, print or online); \$96 (individual, print & online)

Proteomics

editor-in-chief Michael J. Dunn
Wiley-VCH. 12/yr. 518 euros, \$518 (institutional); 98 euros, \$98 (individual)

Samuel Broder

The first complete genomic sequence of a free-living organism, the bacterium *Haemophilus influenzae*, was published in 1995. Since then, many single-cell genomes have been sequenced, and more recently, large multicellular genomes have been completed. The genomes of complex organisms such as the fruitfly, mouse and human are widely available as a resource

for scientific discovery.

Although all of these developments are quite recent, it is already difficult to remember what science was like in the 'pre-genomic' era. Ultimately, understanding these genomes may reveal which human attributes are innate and which are acquired, and how the interplay between heredity, lifestyle and the environment affects our susceptibility to disease.

Such an understanding may eventually allow us to complete the journey towards comprehending how DNA varies among people suffering from common diseases, and in particular, exactly how such variations cause illness and affect our response to drugs. The study of the genome and the associated protein content (the proteome) of free-living organisms may some day make it possible to localize and grasp the function and regulation of every human gene. Alas, we are not there yet. Four journals may be useful to scientists who are pursuing these goals.

Pharmacogenomics provides thoughtful reviews, meeting and conference information, a regular WebWatch section and related material in the rapidly expanding field of pharmacogenomics. The articles cover diverse topics. One excellent article is a review of the scientific opportunities for vaccinating against meningococcal disease using vaccines based on our knowledge of bacterial genomics (and comparative genomics generally). The availability of complete genomic information offers unique resources for identifying virulence factors, new vaccine targets and the mechanisms for possible bacterial 'escape' from host defences (for example, antigenic variation occurring through slipped-strand mispairing).

Other articles range from a discussion of variations between individuals — polymorphisms — in the structure of special transporters of molecules that are important in pharmaceutical uptake and retention to how polymorphisms in other crucial molecules can affect the severity of asthma or drug response in schizophrenia. One need not be an expert to read these articles, and yet there is very useful content even for those who are experts.

Physiological Genomics specializes in taking genomic and related sequence information and correlating it with important physiological parameters, often at the level of the whole animal. Thus, one can find original papers describing the genetics of

differences in pulmonary functional residual capacity in mice, intrinsic aerobic endurance in rats, differences in bone density in various murine models, and the preservation of homeostasis by cooperation between two alternative genotypes of genes involved in the renin-angiotensin system. The articles generally report interesting, original research, and are well written and edited. Over the years, some have argued that whole-organism physiological research may perhaps have been overshadowed by 'pure' molecular biology in the genomic era. This journal is a wonderful antidote to that view.

Current Genomics provides broad, general overviews of topics in structural and functional genomics, including the computational biology necessary to support these fields. There are interesting reviews on

a range of topics, from DNA methylation and breast cancer to mammalian homologues of fruitfly *flightless 1* genes, which are involved in early development. I particularly enjoyed the paper by Florea et al. on validating computer programs for functional genomics in gene regulatory regions. Some of the reviews,

however, may be rather too

general for the specialist reader.

Proteomics provides important original papers in this growing field. A noteworthy finding in the human genome is the relatively low number of protein-coding genes. Before the human genome was sequenced, it had been expected that there would be around 120,000 genes, or possibly more. The actual number is closer to 30,000. Considering that the fruitfly genome contains about 14,000 genes and that of the mustard plant, *Arabidopsis thaliana*, about 26,000, we need to look beyond gene number *per se* in addressing scientific issues on human complexity and targets for novel pharmaceutical agents. Thus, new approaches to studying structural and functional proteomics, and, in particular, modifications that occur after the gene has been translated, are more critical than ever.

Proteomics is targeted at scientists who are actively engaged in proteomics-based research. In-depth and highly technical reports summarize a range of topics, from the effects of gas-phase conformation on peptide fragmentation to methods for improving the accuracy of matrix-assisted laser desorption/ionization-mass spectrometry. Subtle advances in highly technical aspects of proteomics research



are likely to profoundly influence the speed with which the human and mouse genomes can be applied to prevent, diagnose and treat cancer, cardiovascular disease, diabetes and other important illnesses. Such advances are likely to be reported in this journal.

I encourage readers to sample all four journals, and to include them as resources for their research. ■

Samuel Broder is executive vice-president for medical research, Celera Genomics, 45 W. Gude Drive, Rockville, Maryland 20850, USA.

■ www.ashley-pub.com/html/journal_home.asp?JournalID=13

■ www.bentham.org/cg/index2.html

■ www.wiley-vch.de/publish/en/journals

■ www.the-aps.org/publications/journals/pub_pg_home.htm

The Earth cubed

Geochemistry, Geophysics, Geosystems (G^3): An Electronic Journal of the Earth Sciences

editors William White, Henry Elderfield & Richard O'Connell

American Geophysical Union. Access free

Donald E. Canfield

I, for one, greatly enjoy thumbing through my favourite journal. As a research scientist I already spend too much of my time in front of the (stupid) computer, and I am reluctant to concede the joy of the bound hard copy. But despite my old-world attitudes, I am very positively impressed with the new all-electronic journal *Geochemistry, Geophysics, Geosystems (G^3)* from the American Geophysical Union.

The journal serves a broad audience within the geosciences, from biogeochemistry to cosmology, and emphasizes cross-disciplinary approaches to understanding the Earth, and indeed the Solar System, as functioning, evolving, systems. The journal is backed by an extremely strong group of editors and associate editors who, together with the editorial advisory board, represent some of the top researchers in the geosciences.

Papers may be published as Articles, Research Letters (short papers with significant findings), Reviews, Data Briefs

(unpublished data with minimal interpretation), Technical Briefs (methods including computer programs), Characterizations (models with minimal documentation), Commentaries and Editorials. G^3 is clearly striving to maximize the dissemination of scientific information by embracing forms of publication, such as the Data Briefs and Characterizations, that previously had little place in the traditional literature. These formats, however, have not yet been exploited by authors in G^3 .

A particularly innovative aspect of G^3 is the publication of 'special issues' whose papers are posted as they are accepted for publication. Thus, the special issues assemble papers as they become ready, eliminating the frequent frustration surrounding the delayed publication of thematic issues in traditional journals.

As a principal goal, G^3 offers authors numerous potential advantages unique to web-based publishing. These include the publication of large databases, and innovative graphics such as movies, virtual-reality images and sound. Another particularly attractive feature is a short turnaround time from submission to publication. Research Letters can find their way into 'print' within a couple of months, whereas Articles are posted, on average, six months after submission, with about three months between acceptance and publication.

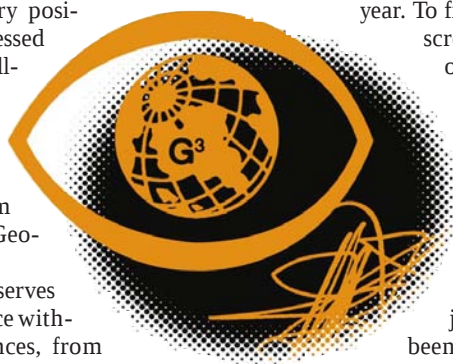
When posted, papers are assigned a volume number, a paper number (including number of words) and a publication year, and this is how the papers are to be referenced. No page numbers are assigned. The current citation format, however, is likely to be revised by the end of the year. To find a paper, one must either

scroll through the posted papers or browse either the author index or the publication-date index. I find this somewhat cumbersome, and a more inter-active search engine is promised in the near future.

All in all, G^3 is an innovative step into the future of journal publishing, and has been embraced by the geosciences community, with a number of first-rate publications already posted. The advantages of rapid publication, low cost (currently, the journal is free of charge) and non-traditional publication formats should ensure the continued growth and expansion of G^3 . I only wish I could put it on my shelf. ■

Donald E. Canfield is at the Institute of Biology, University of Southern Denmark/Odense University, Campusvej 55, 5230 Odense M, Denmark.

■ <http://g-cubed.org>



Towards a common future

Environment, Development and Sustainability: A Multidisciplinary Approach to the Theory and Practice of Sustainable Development

editors Bhaskar Nath, Luc Hens & David Pimentel

Kluwer. 4/yr. 263 euros, \$263 (institutional); 70.50 euros, \$88 (individual)

David R. Brown

Conflicts between development and the environment first significantly permeated the public consciousness in the 1960s, catalysed by such prophetic texts as Rachel Carson's *Silent Spring* and Paul Erhlich's *The Population Bomb*. The world's population was burgeoning, resources were belatedly recognized as exhaustible, and the two following decades recorded a mounting roll-call of environmental degradation — acid rain, ozone depletion, deforestation, loss of biodiversity.

Against this backdrop, in 1987 the World Commission on Environment and Development published its summary report, *Our Common Future*, ushering in the goal of sustainable development, and famously defining it as "development that meets the needs of the present generation without compromising the need of future generations".

A laudable goal, undoubtedly, but where are the goalposts? With more than 100 extant definitions of sustainable development, and countless words written about it, *Environment, Development and Sustainability*, launched in 1999, seeks to make its own contribution. The first issue included papers that returned to fundamental but unanswered questions from the 1960s: is

there enough spare land for food production in developing countries, and will the limits of Earth's resources control human numbers?

The journal's stated aims are exceptionally broad, but are intended to provide a multidisciplinary appraisal of the theory and practice of sustainable development. It achieves this with an interesting range of generally high-quality articles. Its international remit is realized, although with an emphasis on developing countries. Topics addressed to date include environmental and socio-economic aspects of forest land allocation and deforestation in Vietnam, the ecological impacts of Chile's neo-liberal policies and tropical coastal organisms as indicators of mercury pollution in India.

Although such research could be published in a range of outlets, the thread of sustainable development is woven through the papers and the journal stands up well as an integrated whole. This is reinforced by the inclusion of work related to policy (renewed efforts by the Organisation for Economic Co-operation and Development to address sustainable development are documented) and theory. Specific methodologies offering the potential to chart progress towards (or away from) sustainability are addressed, with a robust defence of ecological footprinting—a technique to try to link the land-area resource requirements to per capita consumption for different societies. Training initiatives also fall within the journal's remit, with an account of a simulation model for vulnerability to climate change.

Each issue comprises between four and six papers of variable length (submission to acceptance times are typically six to ten months), followed by a book review. Given the wide range of subjects encompassed, any particular issue of the journal can touch on only a few, but the selections presented are well balanced and will appeal to those wishing to keep abreast of current sustainability-related topics. Special issues look set to become a feature, with the first—"Local knowledge in the tropics: Relevance to conservation and management"—being the proceedings of the annual meeting of the Association for Tropical Biology.

Environment, Development and Sustainability was launched in the last year of a century during which, by any definition, development had become increasingly unsustainable. Whether it will document continued decline or positive progress remains to be seen, but the journal shows the potential to offer a valuable commentary. ■

David R. Brown is at the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

▶ www.wkap.nl/journalhome.htm/1387-585X

Room for one more on board?

Advanced Synthesis & Catalysis

executive editor Joe P. Richmond
Wiley-VCH. 8/yr. 428 euros, \$498
(institutional)

Amir Hoveyda

Catalysis and synthesis are two of the most important areas of investigation in modern organic chemistry. These fields have critical implications for medical and pharmaceutical research as well as for materials science. Not surprisingly, many outstanding, internationally recognized journals dedicate a notable portion of their pages to describing advances in these areas. *Journal of the American Chemical Society* and *Angewandte Chemie* are among at least ten respected publications that regularly cover the latest research in catalysis and synthesis. Such publications deal with novelty of approach and with issues of practicality, scale and environmental friendliness.

It has become an accepted dictum that a new method in synthesis must strive to satisfy some or all of the above criteria; and papers introducing a new catalyst or synthesis method generally include discussions of these issues. After all, some of the most important aspects of catalysis concern issues of economy and the environment. Papers in the journals mentioned above frequently cover attempts to develop recyclable supported catalysts, those that do not require a solvent, or processes that have a high turnover number and/or frequency. And several journals regularly publish symposia-in-print that cover exciting new related topics.

It is therefore difficult to understand immediately what niche *Advanced Catalysis & Synthesis* is expected to fill. The publishers state that the new publication focuses on "chemical reactions that are economical, safe, environmentally benign, resource- and energy-saving". But such attributes have always been part of the yardstick by which papers in catalysis and synthesis are measured. For example, the incisive and thought-provoking review article on catalytic kinetic resolution that appears in the first issue could easily have been accommodated

in *Angewandte Chemie*, a journal produced by the same publisher—one might argue that the latter venue would have allowed the article to reach a wider audience.

Most of the remaining reviews and communications in the journal's first four issues address questions of efficiency and selectivity. A few others are directed more towards the practical aspects of synthesis and catalysis, but all could have been submitted to any of several existing journals.

Perhaps the most striking feature of this new publication is its illustrious editor-in-chief and editorial board; a more respected and better-qualified group would have been difficult to assemble. Such a talented and accomplished editorial ensemble clearly have a valid

reason for supporting the journal. But if there is such a mission, it has unfortunately not yet emerged from the first 400 pages of *Advanced Catalysis & Synthesis*. Furthermore, in thinking about the need for new journals in these important and well-represented sub-disciplines of chemistry, and

considering the truncated budgets and rising costs that face most of our libraries, one is reminded of the words of the great twentieth-century architect, Ludwig Mies van der Rohe, that "less is more". ■

Amir H. Hoveyda is in the Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA.

Forum for the genomic onslaught

GenomeBiology.com

editor Theodora Bloom

BioMed Central. \$300, £190 (library, print only), \$130, £80 (laboratory, print & online), \$120, £75 (individual, print & online); \$95, £60 (individual, online only)

John D. McPherson

The past few years have witnessed an explosion of publicly accessible data on the genomes of a wide variety of organisms. These data sets now include physical maps and DNA sequence and annotation of complete genomes. The analyses of such data sets are increasingly large and complex. To better serve the research community, many journals offer a complementary website providing supplemental material that is inappropriate for a printed format.



These supplemental data include additional details on large-scale analysis methods and, often, access to the data themselves. *GenomeBiology.com* is one of the new breed of journals that publish primarily in a web-based format with a supplemental printed journal profiling selected articles and summarizing or merely indexing others with reference to the website.

The main objective of *GenomeBiology.com* is to provide a forum for articles concerning biology that utilize the wealth of genomic data now available, or in their words, "Biology for the post-genomic era". The website/journal aims to be a guide to these genomic resources and to present the impact of the data on biological research in the form of timely reviews and research articles. An extensive international advisory board with expertise in a wide range of organisms and disciplines is charged with determining the appropriateness of submitted articles for the scope of the journal, which is meant to cover all aspects of genomic research, including molecular and cellular biology, genomics, proteomics, bioinformatics, computational biology, large-scale sequence analysis, comparative biology and molecular evolution. The journal succeeds very well in providing a rich resource for publication in these data-intensive research areas, which many traditional journals cannot readily support.

GenomeBiology.com aims for speed of publication, encouraging the use of a direct, web-based upload submission system and a promise of quick review. Although there are no length limits for the web-based articles, most remain succinct and understandable. The primary use of the website as

a publication forum allows the ready inclusion of greater numbers of more complex images and also supports movies and access to raw data.

An additional component of the website is a depository for non-refereed articles with no access restrictions for the research community. Authors are solely responsible for the content of these articles; however, the articles are screened to ensure their relevance to the scope of the journal and to prevent abuse of this service.

Although it is premature to declare this the post-genomic era, *GenomeBiology.com* recognizes the need for journals to adapt to today's data-intensive biological research. Web support for journals, and indeed web-based journals, are essential for providing comprehensive presentation of research material. *GenomeBiology.com* offers research articles of broad interest and insightful reviews. It is an excellent guide to the overwhelming onslaught of genomic and proteomic data and will flourish if its current trend continues. ■

John D. McPherson is in the Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63108, USA.

► <http://genomebiology.com>

On the high side of life

High Altitude Medicine & Biology

editor-in-chief John B. West

Mary Ann Liebert. 4/yr. \$300 (institutional), \$191 (individual)

Ronald J. White and Jancy C. McPhee

Over the past 20 years, interest in the high-altitude life sciences has grown steadily and become increasingly diverse and interdisciplinary. An idea of this can be gained by considering the range of topics — from interleukin-6 levels in high-altitude pulmonary oedema to evolutionary physiology — that are discussed at the International Hypoxia Symposia and other focused scientific meetings.

This growing interest has led to the creation of *High Altitude Medicine & Biology*, a journal with an intentionally broad scope that serves the varied needs of this specialized medical and scientific community. Like the field, the editorial board is diverse, both in expertise and nationality, with world-class members from 17 countries. This speciality journal is off to a good start.

The journal's scope includes high-altitude diseases, physiology, pathology, comparative biology, anthropology, evolutionary biology, and human and animal ecology. This breadth is reflected in the articles in the first few issues, which range



from fundamental gene-based studies to ethnobotany. In addition, the journal hopes to deal with such interesting issues as the practical problems in medicine that develop as more and more people commute to work from low elevations to high altitude.

High Altitude Medicine & Biology does not aim to compete with well-established discipline-based journals, but rather to provide a home for multidisciplinary work, or for articles that do not fit within the scope of existing journals. This policy is wise but presents its own challenge — that of drawing the highest-quality manuscripts in this eclectic field. The editors hope that the journal will actually spur research in the diverse areas within its scope. Only time will tell whether they succeed in this task while maintaining the current quality.

This is an attractive journal, with high-quality print, clear figures, good photographs and a nice layout. Published manuscripts are reviewed and accepted in a reasonable time — about two months — and publication currently follows acceptance rapidly. The articles average about eight pages in length.

A special feature is a section entitled "Sightings", which gives readers interesting glimpses of articles from journals that may be far from their normal reading list. Another feature, "High Altitude Web", provides information on where to look on the Internet for useful information. The journal has space for letters to the editor and book reviews. Interesting case studies, reviews, historical articles and intermittent poetry are all included. Occasionally, abstracts from symposia (such as the International Hypoxia Symposia) are planned.

The journal has already met with some success. It is being indexed by Index Medicus and citations are available through Medline. Additionally, it has been selected as the official journal of the International Society



of Mountain Medicine. If the promise of the first few issues is borne out, this journal could become the standard home for interdisciplinary papers in the high-altitude life sciences. ■

Ronald J. White and Jancy C. McPhee are at the National Space Biomedical Research Institute and Baylor College of Medicine, One Baylor Plaza, NA-425, Houston, Texas 77030, USA.

♦ www.liebertpub.com/ham

Outlook on a crisis discipline

Conservation Genetics

editor A. R. Hoelzel

Kluwer. 4/yr. 217 euros, \$218 (institutional); 63.50 euros, \$60 (individual)

Oscar E. Gaggiotti

As the editor of *Conservation Genetics* notes in the journal's first issue, conservation biology is a field that requires the integration of many disciplines — including social and economic matters — in order to address an issue that concerns us all. But is it then advisable to publish a conservation journal that focuses on only one of the many disciplines covered by the field? The editor, however, puts a strong case for the journal's publication, and stresses that its primary aim is to promote the application of molecular-genetics methodologies to answer pressing questions about the conservation of biodiversity.

The journal publishes research articles, review articles, commentary, short communications and technical notes. It thereby provides a wide range of options for promoting the dissemination of knowledge and for fostering discussion. And this is essential in a field where the results of scientific research can have a direct impact on the long-term survival of species.

As the founder of the field Michael Soulé rightly put it, conservation biology is a crisis discipline, where scientists have to work quickly to address urgent problems. Unfortunately, when operating in crisis mode, there may be a tendency to avoid thinking about new ways of solving the problem at hand and instead to use traditional approaches. Although this is sometimes necessary, it is not a good long-term strategy, and we need a forum for discussing and thoroughly testing new ideas and approaches.

Conservation Genetics could

potentially fulfil this need, but only if it avoids becoming a repository of 'dull but worthy' papers. Purely descriptive papers providing useful information on the conservation of a species or population — for example, phylogeography, genetic structure and level of genetic variability — have been accepted by the journal but represent a small fraction of all the studies it has published so far.

Most of its papers go beyond a simple enumeration of genetic attributes and try to identify possible demographic, environmental and anthropogenic factors responsible for the observed patterns. Other studies have addressed more general and important questions, such as the validity of using neutral genetic markers — those not subject to selective forces — as a surrogate for the additive genetic variation underlying the traits that are the target of selection.

But one cannot help noticing that, despite the very long list of topics the editor suggests as appropriate material for his journal, close to 50% of the papers published deal with the genetic structure and/or phylogeography of populations. This criticism may seem unfair, as it is a trend that has been observed in the early days of other journals covering the broad field of molecular ecology. However, the field of conservation genetics must not be restricted to this narrow area, and the journal should endeavour to attract more papers dealing with other topics, such as inbreeding depression, hybridization and the impact of breeding programmes on genetic diversity.

Despite this caveat, it is clear that the journal has started well, and could eventually become the leading forum for the discussion of all aspects of population genetics in conservation science. ■

Oscar E. Gaggiotti is in the Metapopulation Research Group, Department of Ecology and Systematics, PO Box 17 (Arkadiankatu 7), FIN-00014 University of Helsinki, Finland.



The organic chemist's newswire

Organic Letters

editor Amos B. Smith III

American Chemical Society. Biweekly. \$2,926 (institutional, online & print), \$2,560 (institutional, online only), \$25 (individual, online, ACS members only), \$154 (individual, print, ACS members only)

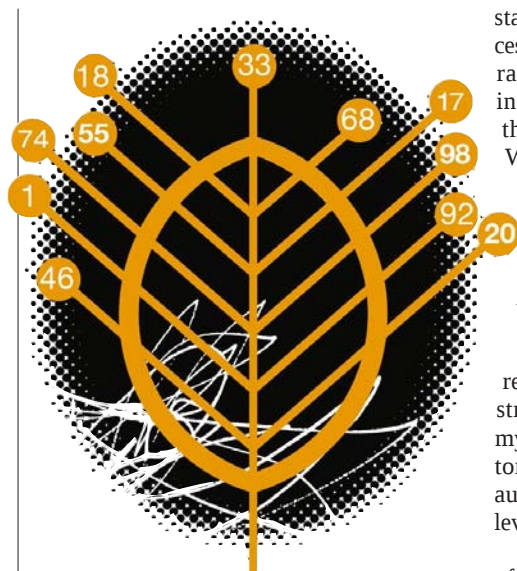
Richard E. Taylor

Interdisciplinary research has led to the creation of new fields and with them a plethora of new journals to collate material within each discipline. One new journal, not based on a new field of science, has taken advantage of another area of remarkable development in the past decade, the Internet. *Organic Letters* is a biweekly journal that offers two- to four-page peer-reviewed "letters" and whose goal is to provide reports on cutting-edge research in a broad range of organic chemistry, including bioorganic chemistry, physical organic chemistry and synthesis. Most importantly, this innovative new journal offers a completely online system of manuscript submission, peer review, editing and publishing.

Organic Letters was established two years ago to fill an important gap in the range of print journals published by the American Chemical Society (ACS). Before *Organic Letters*, organic chemists were limited in their choice of venue for the publication of communications. The ACS journals *Journal of the American Chemical Society* and *Journal of Organic Chemistry* both included "communications" sections, but these were restricted to good-quality manuscripts of broad interest, with experimental details included as web-accessed supporting information. Thus, rapid publication of preliminary communications was often directed towards other "letters"-type journals such as *Tetrahedron Letters* or *Synlett*, non-ACS journals.

With the establishment of *Organic Letters*, the *Journal of Organic Chemistry* stopped publishing "communications". Although it makes perfect sense not to have two society journals competing for quality manuscripts, the consequences are less than ideal. The high-quality manuscripts previously published in *Journal of Organic Chemistry* as communications are now lumped together in *Organic Letters* with many of these more preliminary reports. Despite this negative, the overall quality of the content of *Organic Letters*, in all areas of organic chemistry, has remained at a constant high level.

The web edition of the journal publishes articles within 48 hours of peer-review acceptance under the ACS's innovative "As Soon As Publishable" (ASAP) system. As an



author of several *Organic Letters* papers, as well as a frequent reviewer, I can say from experience that the entire process is remarkably efficient and a model for the future of scientific publication. In addition, the web-based journal, as with all of the society's library, is a powerful reference source. Launched with *Organic Letters* was the Chemical Abstract Service (CAS) Reference Linking feature, in which virtually every reference in the electronic journal is 'hot-linked' to the corresponding CAS abstract, regardless of publisher. Moreover, references to ACS journals have a direct link to the corresponding manuscript.

The remarkably rapid turn-around time from submission to publication gives the reader the feeling of a newswire, reporting the exciting discoveries of the field almost as they occur. *Organic Letters* has a bright future and, from the diverse manuscripts published so far, so does the field of organic chemistry. ■

Richard E. Taylor is in the Department of Chemistry & Biochemistry, University of Notre Dame, 251 Nieuwland Science Hall, Notre Dame, Indiana 46556-5670, USA.

A European slant on mol biol

EMBO Reports

senior editor Frank Gannon
Oxford University Press/European Molecular Biology Organization. 12/yr. £320, \$522 (institutional, print & online); £70, \$116 (individual)

Hidde Ploegh

EMBO Reports is the second periodical to be produced jointly by Oxford University Press and the European Molecular Biology Organization. *EMBO Journal*, an enterprise

started more than 15 years ago, is now so successful that not only is it among the highest-ranked molecular-biology journals, but its income also helps EMBO pay for useful things such as postdoctoral fellowships. Will its sibling, *EMBO Reports*, be equally successful?

The launch issue sets forth the aspirations of *EMBO Reports*. As far as the science is concerned: shorter articles than *EMBO Journal* accommodates, but of the same broad scope and high quality. Here I can be brief. There is a good amount of really exciting science packed into the stream of *EMBO Reports* steadily arriving on my desk over the past year and a half. The editors achieved their goal by roping in the authors necessary to produce at the same level as *EMBO Journal*.

By far the most interesting new aspect of the journal is something not found in *EMBO Journal*: journalism in the form of editorials, interviews, book reviews and policy pieces. The meeting reports have substance and are far more informative than those I read in competing publications. The news editor, Holger Breithaupt, does an excellent job in several lengthy interviews — with, for example, Kai Simons, Harold Varmus, Noel Treacy, Ralf Pettersson — and commentaries.

EMBO Reports cannot deal with the truly topical issues in the manner of a *Science* or *Nature*, simply because of its publication schedule, but — to name just two examples — I found the timely articles on the human genome and reproductive issues and those on the stem-cell debate thoughtful, interesting and highly readable. Thankfully, the publishers care more about substance than page limits. Also, scientists surely enjoy reading about scientific gossip and human-interest issues — the Folkman–Abbott lawsuit, some of the inner workings of the Nobel committee, concerns about degree requirements.

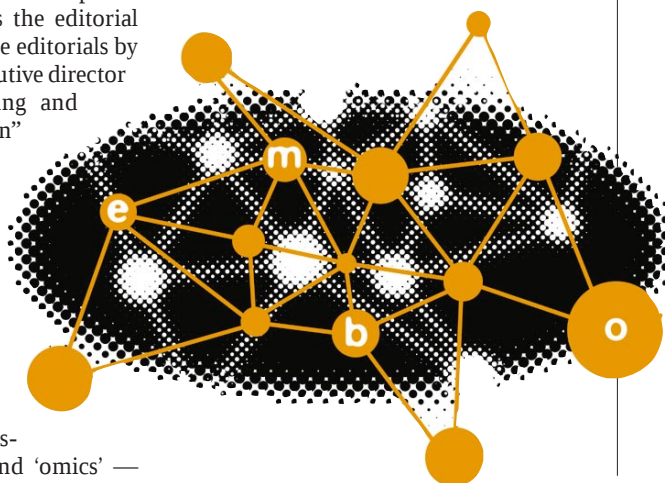
Of the major science publications, only a few avail themselves of the opportunity to dish out strong medicine in the opinion and policy arena, and I hope that future pieces will pack as much punch as the editorial board have so far allowed. The editorials by senior editor and EMBO executive director Frank Gannon are interesting and provocative ("Back to Darwin" is the best of the lot), although somewhat variable in octane number. An appeal for science and our role as scientists to be portrayed in a positive manner is voiced by many other publications, and I wonder what such preaching to the converted will amount to. But then, there is also a biting assessment of the fad of 'omes' and 'omics' —

transcriptome, glycosylomics; even a lowly 2D-gel apparatus will henceforth be a proteomics facility. Or the question: will hypothesis-driven research be relegated to the same plane as the flat-Earthers? This insane opinion, anonymously cited by Frank Gannon, came from a "witless 'genomics' drone somewhere in Germany" — attribution of this quote would have served more than a journalistic purpose.

I tend to think of science as a fairly global enterprise, but a 'European' odour emanates from some of the news analyses. And I wish that Europe's scientists would tell Brussels what they want and need, rather than exist in reactive mode — 'frameworks' and Eurocrats be damned. *EMBO Reports* could be a voice for a serious critique of Brussels' bureaucratic bungling, constructive exchanges with Euro-commissioner for research Philippe Busquin notwithstanding. Even if it is funded with money from member states of the European Commission, a watchdog function for *EMBO Reports* might be on the cards, hopefully without fear of budgetary retribution.

The journal is a pleasure to leaf through: glossy paper, generally outstanding reproductions of photographs, high-quality graphics, all in a sober, almost conservative, design. A journal can no longer distinguish itself by showing on its covers the usual immunofluorescent and crystallographic images or their abstractions. Why not steal an idea from *EMBO Journal* and on occasion have real cover art? Spice up the typographic design, and commission (European) designers and artists to lend their voice to our science.

Along similar lines, a likeable feature is the inclusion of photos of the authors of reviews, interviews, meeting reports and the like. This raises the question: why not have pictures of *all* the women and men behind the work? *EMBO Reports* is a winner: if it continues to publish excellent science and grows editorial/opinion fangs of appropriate size, it will enliven debate and render an exceptional service to molecular biology. It is wonderful to see a new



publication take on journalistic responsibilities along with the science. ■

Hidde Ploegh is in the Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

♦ www.embo-reports.oupjournals.org

The way of the future?

BioMed Central (online only)

editorial directors Peter Newmark (biology) & Fiona Godlee (medicine)

BioMed Central. Access free

William Hersh

There is a growing concern among scientists that research results are controlled by an increasingly small number of publishers who have great control over the marketplace. This monopoly, however, is indirectly supported by scientists themselves, who aim to publish in the most prestigious and well-cited journals and need those publications to obtain promotion and grant funding. A related concern is that governments fund most basic research, individual scientists perform it, and then reports of the results are handed over to publishers who assume the copyright and sell it back to those very governments and scientists. Because many scientific journals are costly, accessing them is increasingly beyond the reach of both individuals and libraries, the latter of whom are increasingly strapped financially.

These concerns have been addressed by a number of individuals and organizations. In 1999, the US National Institutes of Health (NIH) announced an initiative, E-Biomed, later renamed PubMed Central (PMC), which would provide a repository of scientific research papers that could be accessed free over the World Wide Web. The proposal met with much controversy, especially the part, later dropped, that proposed an additional repository of papers before peer review.

After considerable debate — and some retrenching on the part of NIH — PMC has gone live and currently has about a dozen journals available (www.pubmedcentral.gov). One restriction whose relaxation has led to more journals joining the PMC initiative has been the abandonment of the requirement for articles to be physically located on the PMC site. Now journal publishers need only provide links to PMC and keep their papers on their own sites. Some view the PMC model as an extension of the approach used by GenBank, where data from genomics research are made widely accessible.

Another initiative to make research articles more widely and freely available is the

Public Library of Science (PLS; www.publibraryofscience.org): scientists who sign the PLS open letter pledge only to publish in and subscribe to those journals that agree to make their research reports freely available within six months of publication on the web via PMC or any other equivalent model. More than 28,000 scientists in 172 countries have currently signed the letter, agreeing to adhere to the pledge from this September. These initiatives have been debated vigorously in many forums, including the *Nature* website (www.nature.com/nature/debates/e-access/index.html).

A new approach to scientific publishing that is completely electronic and incorporates the goals of PMC and PLS has now been launched by BioMed Central (BMC). This is an initiative based in the United Kingdom to provide “peer reviewed research across all areas of biology and medicine, with immediate, barrier-free access for all”. Of course, such a goal requires a business model, and BMC is currently supported through advertising on its site. It plans, in the future, to charge authors for processing and editing of their papers, although this charge will be waived for those unable to afford it. Ultimately, the BMC model views the costs of publishing, which are greatly reduced in the electronic model, as a cost of research to be borne by those who fund it, such as government agencies and commercial concerns.

BMC currently houses 18 biology and 39 medicine journals, each named and devoted to specific subject areas, for example, *BMC Bioinformatics* and *BMC Physiology*. The site also contains four affiliated journals as well as abstracts from a small number of scientific meetings. Articles are published in native HTML web-page format as well as a more visually pleasing and better-printing PDF format. A relatively simple search engine allows searching of individual journals or all titles across the site. The search engine also allows queries to be passed to PubMed or PMC. All publications in BMC, since its inception, have been part of PMC and meet the criteria of the PLS.

The BMC publishing process has many aspects that will appeal to scientists. First, the peer-review process is rapid — the website claims that the time from submission to publication averages 35 days. This is due to the use of a completely electronic process, including online submission and peer review. In addition, because all publishing is electronic, articles can be published as soon as they are accepted.

The BMC website also boasts other

advantages to its process, including the lack of space constraints, which means that papers worthy of publication do not need to compete with one another for finite space as in print journals. Authors also maintain copyright of their articles, although they grant BMC an exclusive licence to republish the article, even in print form. As all BMC articles are indexed in Medline, research published in BMC is no more difficult to find than that published elsewhere. BMC also provides a standard means for citing articles, which should ensure that they are easy to cite as well as access. Finally, the organization is working with a number of large-scale archiving efforts, such as the Open Archives Initiative (www.openarchives.org), to ensure that content is accessible in perpetuity.

There is concern, of course, that articles published in BMC may not have the ‘prestige’ associated with traditional scientific publishing. Certainly, there is no inherent reason why high-quality science cannot be published under this model. To assess the quality of science in BMC, I suggest that each reader assess the BMC

website articles from his or her field. I can say that articles from

my own field

(medical informatics) are all

of good quality,

even though the

most cutting-edge

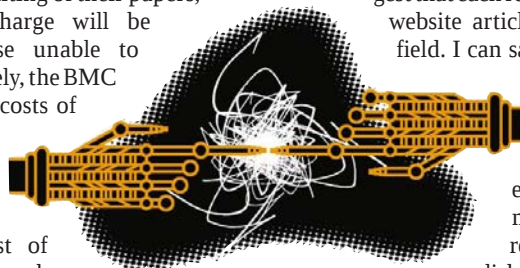
research is still published in the traditional

print journals. But this will change if more top-notch research is submitted to BMC and efforts like it. I have served as a peer reviewer for BMC and can state that I applied my usual rigour to the process.

I believe that BMC represents the future of scientific publishing, and that such a future has great potential. In this era of electronic publishing and growing concern over access to research papers, the BMC model represents a credible alternative. Its success and that of similar efforts depend upon we scientists and our decisions to submit articles to BMC and to cite those it has published. To its credit, BMC does not shrink from the notion that there are costs to scientific publishing, even though they can be reduced by using electronic processes. For BMC and similar models to succeed, those who fund research must realize that the modest cost of publication is a reasonable one that should be associated with research. ■

William Hersh is in the Division of Medical Informatics & Outcomes Research, School of Medicine, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, BICC, Portland, Oregon 97201, USA.

♦ www.biomedcentral.com



In the name of love?

By defining ideas precisely, as science does, fiction would deny its readers freedom of interpretation.

Alan Lightman

During my career I have had the privilege of working both as a physicist and as a novelist. As a member of both of these communities, I've been fascinated by the different ways in which they work, the different ways in which they think, and their different approaches to truth.

One important distinction that can be made between physicists and novelists, and between the scientific and artistic communities in general, is in what I shall call 'naming'. Roughly speaking, the scientist tries to name things and the artist tries to avoid naming things.

To name a thing, one needs to have gathered it, distilled and purified it, and attempted to identify it with clarity and precision. Consider, for example, the word 'electron'. As far as we know, all of the zillions of electrons in the Universe are identical. There is only a single kind of electron. And to a modern physicist, the word 'electron' represents a particular equation — the Dirac equation with field operators.

That equation summarizes, in precise mathematical and quantitative terms, everything we know about electrons — every interaction, the precise deflections and twists of electrons by particular magnetic and electric fields, the tiny effects of electrons and their antiparticles materializing out of nothing and then disappearing again. In a real sense, the name 'electron' refers to the Dirac equation. For scientists, it is a great comfort, a feeling of power, a sense of control, to be able to name things in this way.

The objects and concepts with which the novelist is concerned cannot be named. The novelist might use the words 'love' and 'fear', but these names do not summarize or

convey much to the reader. For one thing, there are a thousand different kinds of love. There's the love you feel for a mother who writes to you every day during your first month away from home, and the love you feel for a mother who, when you stumble into the house drunk after driving home from the prom, slaps you and then embraces you. There's the love you feel for a man or a woman you've just made love to, and there's the love you feel for a friend who calls to give you support after you've just split up with your spouse. But it's not just the fact that there are so many different kinds of love that prevent the novelist from truly naming the thing. It's also the fact that the idea of love — the particular sensation out of the thousands of different kinds of love — must be shown to the reader not by naming it, but through the actions of the characters.

If love is shown, rather than named, each reader will experience it and, what's more, will understand it in his or her own way. Each reader will draw on their own adventures and misadventures with love. Every electron is identical, but every love is different.

The novelist doesn't want to eliminate these differences, doesn't want to clarify and distill the meaning of love so that there is only a single meaning, like the Dirac equation, because no such distillation could represent love. Any attempt at such a distillation would undermine the authenticity of readers' reactions, destroying the delicate, participatory creative experience of a good reader reading a good book. In a sense, a novel is not complete until it has been read. And each reader completes the novel in a different way.

I'll give another illustration of the difference between naming and not naming. Let me represent science by expository writing. Like science, a piece of expository writing

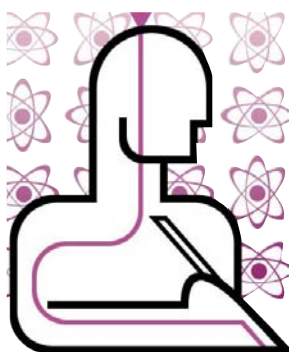


takes a reductionist and reasoned approach to the world. You have a position or argument, you structure this argument in logical steps, amassing facts and evidence to convince your reader of each assertion. We all learn that in expository writing it is useful to begin each paragraph with a topic sentence. A topic sentence, in effect, states the idea of the paragraph at the outset. You thus begin by telling your reader what he or she is going to learn in the paragraph and how to organize his or her thoughts so as to gain as ordered and structured an understanding as possible.

But in fiction, a topic sentence is usually fatal, because the power of fiction is emotional and sensual. You want your reader to feel what you're saying, to smell it and hear it, to be part of the scene you are creating. You want your reader to be blind-sided, to let go and be carried off to a magical place. Every reader will travel differently, depending on his or her own experiences of life. With a topic sentence, you don't leave room for your reader's own imagination and creativity to be engaged as the paragraph unfolds. The difference can be stated in terms of the body. In expository writing, you want to get to your reader's brain. In creative writing, you want to bypass the brain and go straight for the stomach or the heart.

Alan Lightman is adjunct professor of humanities at the Massachusetts Institute of Technology and the author of *Einstein's Dreams* and *The Diagnosis*.

Adapted from "The physicist as novelist" in *The Future of Spacetime*, edited by Richard H. Price (W. W. Norton, New York, in the press).



In fiction, you want your reader to feel what you're saying, to smell it and hear it, to be part of the scene. You want your reader to be blind-sided, to let go and be carried off to a magical place.

An intriguing door

Ian Glynn

The idea that the brain is the seat of the mind has been around for two-and-a-half millennia, but the phrase “neural correlates of consciousness” has appeared only in the past few years. Its popularity (Francis Crick recommends ‘neural correlate’ to his readers as a “useful term ... worth committing to memory”) is sometimes attributed to its neutrality — correlation between neural and mental events neither implies nor excludes a causal relationship between them.

So the phrase can be used happily by old-fashioned interactionist dualists (which is, of course, what most of us are most of the time) as well as by physicalists, functionalists, epiphenomenalists, identity theorists and what have you (each of which is what some of us are some of the time). Even behaviourists, who ignore consciousness, and eliminative materialists, who dismiss it as a misleading term in folk psychology, cannot deny that neural activity accompanies what they ignore or dismiss. But all this is much too

cynical. The neural correlates of consciousness are fashionable because they are important and almost accessible.

Ignoring tall stories, we know that consciousness is always associated with neural activity in a complex brain, and that altering that activity, whether by changes in sensory input, injury, disease, drugs, direct electrical stimulation or neurosurgery, can alter the contents of consciousness. So a reasonable hypothesis is that individual conscious states exist only in association with particular patterns of neural activity, and that the existence of these patterns is always associated with the corresponding conscious states. It is these patterns that are referred to as the neural correlates of consciousness.

Why such patterns should be associated with thoughts or feelings is among the most difficult of all problems. When Charles Darwin asked “why is thought being a secretion of the brain more wonderful than gravity a property of matter?” he was not minimizing the difficulty but rather arguing that failure to understand how the brain produces thoughts should not lead us to reject the idea that it does. Difficult though the problem is, determining what the neural correlates of consciousness are would seem to be a possible first step towards solving it. The trouble is that taking even this first step involves simultaneous monitoring of two elusive sets of events: the contents of consciousness and the relevant neural activity in a vast collection of neurons that are involved in many different processes.

During the past 12 years, ingenious experimenters in several laboratories have attempted this simultaneous monitoring. We now know a good deal about the earlier parts of the neural-event chains that link stimulation of sense organs to the resulting sensations or perceptions. Therefore, a general strategy is to follow the chain of events initiated by some sensation- or perception-causing stimulus in the hope that we will find a pattern of neural events that not only correlates well with the particular conscious experience but can also be shown to be both necessary and sufficient for it to occur.

That goal has not been reached, but in one of the most striking experiments so far, Nikos Logothetis and colleagues exposed a monkey to a visual stimulus that was ambiguous and could give rise to two different conscious perceptions. Having trained the monkey to indicate which perception it currently had and when the perception switched between alternatives, the researchers looked for (and found) neurons in the visual areas of the monkey’s brain that altered their behaviour when the

Consciousness

Can we find patterns of neural events that are both necessary and sufficient for particular conscious experiences to occur?

perception changed. The logic of this procedure is that the monkey signals what it thinks it sees rather than what it is looking at. Because the stimulus remains constant, all of the early processing should remain constant, and the neurons that change their behaviour when the perception changes are therefore suspected of being specifically involved in conscious perception. This is still a long way from our goal but it is, as Ilya Farber and Patricia Churchland put it, “a substantial foot in a most intriguing door”.

If individual conscious states and processes are inseparable from particular patterns of neural activity, they can be effective in the physical world, and the evolution of consciousness through natural selection is plausible. But in 1879 William James pointed out a problem. There is a strong correlation between the pleasantness or unpleasantness of sensations and the survival value of the activities or situations that cause them. We enjoy eating, drinking and sex; we dislike hunger, injury and exhaustion. As long as we believe in old-fashioned, common-sense, interactive dualism, this correlation is predictable; our subjective mental states influence our behaviour, so, for our ancestors to have survived, their subjective preferences must have tended to favour their survival.

But once we give up this belief in interactive dualism (as most of us do when we think in ‘neuroscience mode’) and assume that it is only through their neural correlates that mental states can influence behaviour, the subjective character of mental states has no relevance. Why, then, is there an association between pleasure and survival-promoting situations, and between discomfort and survival-threatening situations? It seems that there is a stumbling block on the threshold of the intriguing door. ■

Ian Glynn is at Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK

FURTHER READING

Glynn, I. *An Anatomy of Thought: The Origin and Machinery of the Mind* (Weidenfeld & Nicolson, London; Oxford Univ. Press, New York; 1999).
Logothetis, N. K. *Sci. Am.* November 44–51 (1999).
Richards, R. J. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior* (Univ. Chicago Press, Chicago, 1987).
Crick, F. *The Astonishing Hypothesis* (Scribners, New York, 1994).



That thinking feeling: different conscious states are thought to have specific neural patterns.

Fat and formation in flight

Jeremy M. V. Rayner

Predictions of flight performance in birds rely heavily on aerodynamic theory because it is difficult to measure energy consumption in flight. Fresh data leave part of the theory up in the air.

Active flight in birds is one of the most demanding forms of locomotion, and imposes so many constraints on bird evolution that few aspects of bird biology are untouched by it. The usual currency used to assess flight performance is flight power, which is the rate at which the bird consumes metabolic energy and generates mechanical work as it flies. Ever since birds inspired aircraft designers in the nineteenth century, there have been numerous attempts to measure flight power. But there is still debate about how best to estimate it, and in particular about whether estimates based on aerodynamic theory are reliable enough to predict how much energy a bird uses in flight.

Measuring the total energy consumption in a free-flying bird is technically demanding. So instead, much effort has been invested in developing theories of power consumption by modelling the aerodynamics of flapping flight, and adding plausible assumptions about the efficiency of converting fuel into mechanical work and about heat generation by the flight muscles. Calculations of the aerodynamic portion of power output, derived from modelling the forces generated by the bird's wings, appear to be relatively accurate¹. But tests of the predicted total power, including the heat generated, have been equivocal^{1,2}. Papers in this issue describe the use of total power measurements to test two long-cherished aerodynamic predictions. Kvist and colleagues³ (page 730) investigate how power varies with body mass in individual birds, and Weimerskirch *et al.*⁴ (page 697) look at the effects of flying in formation on energy expenditure.

Kvist *et al.*³ quantified variation in flight power with body mass by using radioactively labelled isotopes to measure the oxygen metabolism of birds flying in a wind tunnel. The wind tunnel allows the observer to control the flight speed and other conditions; the bird thinks it is flying forwards but, because it is not moving relative to the observer, it is comparatively easy to measure wing movements or flight physiology.

The birds involved were red knots (Fig. 1), which are relatively small wading birds that can fly enormous distances during migration. Like many migrants, knots increase their body mass before departure by



Figure 1 On the wing — red knots, the birds Kvist *et al.*³ worked with. The photograph above shows the birds flying in a wind tunnel.

depositing fat. The fat forms the fuel the bird will consume during its flight, and is a substantial additional load that can sometimes double the body mass. The total mass of a bird is a major constraint on its flight mechanics, but little is known about the biomechanical compensations a bird must make to cope with a changing fuel load. Knowing how changing mass affects flight power, and how fuel load, mass and power decline in the course of a long flight, are therefore crucial to understanding some of the decisions that birds make in flight⁵ and the ways in which their body mass varies.

Mathematical predictions about how flight speed and power vary with load go back to the use of balloons during the siege of Paris in the Franco-Prussian war of 1870–71 (ref. 6). But the theory is equally applicable to aircraft and birds⁷, because the wings of most flying animals and all but the fastest aircraft work in the same way. Two assumptions go into this prediction. The first is that bird shape is much the same across species⁸, so the bigger the bird, the bigger its wings. The second is that all birds fly with the same lift coefficient — that is, the wings of different birds generate the same lift force, allowing for changes in wing area and speed with size.

The mathematical outcome of the theory is that heavier birds have to fly faster, and that flight speed should vary with the total body

mass M as $M^{1/6}$. Accordingly, power — which is proportional to the product of weight and speed — scales as $M^1 \times M^{1/6} = M^{7/6}$. These widely quoted^{7,8} formulae are the foundation of the aerodynamic estimation of bird flight power^{1,8}. But they are by no means robust: although it is difficult to test the speed prediction, measured power seems to scale less steeply, as $M^{0.78}$ (ref. 2).

However, these formulae are only part of the story for the knots. Kvist *et al.*³ worked with four individual birds. In each of these, total mass including the fuel load varied naturally by up to 75%, but the wings of course remained the same size. The theory needs to be recalculated to allow for this; if the lift coefficient is again constant, then speed should vary as $M^{1/2}$, and power as $M^{3/2}$. Little is known of the mechanical constraints on how a bird chooses or controls speed, so Kvist and his colleagues flew all the birds at a single representative speed in their wind tunnel. In this case the rise of power with mass is predicted to be slightly less steep, in the region of $M^{1.25}$ according to simulations with a widely used model of flight power⁷, but the scaling index is so large that a matching change in total power should be obvious in the measurements. Although power does increase with mass, the prediction fails dramatically: the measured³ scaling $M^{0.35}$ is far shallower than expected. Heavy birds use less power than they should.

The implications of this result for both migration and flight mechanics may be

A. LINDSTRÖM

considerable. The slow increase of power with mass means that large fuel loads can be carried comparatively cheaply. So mass dynamics in migrants are not determined simply by the costs of flight, but must also reflect the benefit of having sufficient fuel reserves to compensate for vagaries of the weather or other uncertainties. This makes it easier to understand how the extraordinary flight achievements of migrants might have evolved. In the case of the knot, for instance, it helps to explain how a small bird with a lean mass of only 100 grams can fly up to 5,000 km from Britain to the Russian Arctic in a single hop.

The immediate conclusion is that aerodynamic flight-power models (see refs 1 and 7, for example) do not predict total power reliably under all conditions. One possible explanation is that a bird with a full fuel load has some biomechanical means of compensating for the rise in aerodynamic power needed to support its weight, perhaps by manipulating its wing lift coefficient or by varying the rate at which it moves its wings. But it is far from obvious that any such mechanism could account for the discrepancy in power. Rather, Kvist *et al.*³ suggest that the efficiency with which muscles convert fuel to mechanical work also rises with load for individual birds. This proposition is consistent with a calculated rise in flight efficiency with body mass between birds of different sizes^{1,2}, and with an increase in efficiency with flight speed in starlings². But it also raises a paradox: why can't a bird carrying a small fuel load operate at the high efficiency that it can reach when carrying a large load?

The situation with fuel load may be complex, but the implications of Weimerskirch and co-workers' observations⁴ on the benefits of formation flight are much clearer. Reassuringly, in this case the predictions based on aerodynamics do not fail. The idea that flying in formation gives significant aerodynamic benefit dates back to a 1914 paper by Wieselsberger⁹, at the time a doctoral student under Ludwig Prandtl at Göttingen. The date is significant: the paper followed soon after Prandtl's 'lifting-line' principle of wing action of the same year, which described how vortices around the wing and in its wake are responsible for aerodynamic lift, and as such was one of the fundamental results of modern aerodynamics. The wake vortices force air downwards in the region behind the wing, but air is forced upwards outside the wake. Wieselsberger realized that this principle should apply to birds just as it did to aircraft, and that a group of birds could exploit the updraft to fly more cheaply if they adopted a V-shaped formation in which each bird flies in the up-current generated by the one in front.

Not all ornithologists have accepted the aerodynamic explanation for this spectacular

flight phenomenon, however, and formation flight has been the subject of an unresolved debate between the 'aerodynamic' and 'behavioural' camps. Until now, neither side of the argument has been complete. On one hand, the aerodynamic prediction has never been fully quantified because the magnitude of the energy saving depends on the group's geometry, and the effects of flapping flight or wingbeat synchronization have not been properly modelled. On the other, there are probably benefits from travelling in a cohesive, structured group, with a dominant or experienced leading bird.

In their experiments, Weimerskirch *et al.*⁴ used heart rate in great white pelicans as a proxy for energy expenditure¹⁰. The pelicans were trained to fly after a motor boat and light aircraft, from which measurements and observations were made. The authors found that pelicans in formation had a lower heart rate and wingbeat frequency, and glided more often. These are indications that the birds required less mechanical and total power to fly, so — qualitatively, at least — confirm the aerodynamic prediction. It is likely that aerodynamic and social benefits coevolved to establish this common flight behaviour in large birds.

Taken together, these two sets of measurements^{3,4} add considerably to what we can say about the energetics of bird flight. Aerody-

namic models are, in general, quite good at predicting the biomechanical aspects of flight, but they cannot yet be extended reliably to predict total flight power. We know far too little of the internal physiological processes in flight, and in particular about the efficiency with which the flight muscles generate mechanical work. It is not yet possible to predict this efficiency reliably, or to explain how it varies with mass or speed. If we are to understand the energetics of flight in any bird under any conditions, or are to be able to use simple aerodynamic models to predict total flight power, we will first have to know a lot more about how flight muscle works. ■

Jeremy M. V. Rayner is in the School of Biology, University of Leeds, Leeds LS2 9JT, UK.

e-mail: j.m.v.rayner@leeds.ac.uk

1. Rayner, J. M. V. *J. Exp. Biol.* **202**, 3449–3461 (1999).
2. Ward, S. *et al. J. Exp. Biol.* **204** (in the press).
3. Kvist, A., Lindström, Å., Green, M., Piersma, T. & Visser, G. H. *Nature* **413**, 730–732 (2001).
4. Weimerskirch, H., Martin, J., Clerquin, Y. & Jiraskova, S. *Nature* **413**, 697–698 (2001).
5. Hedenström, A. & Ålerstam, T. *Phil. Trans. R. Soc. Lond. B* **348**, 471–487 (1995).
6. von Helmholtz, H. *Mber. k. Akad. Wiss. Berl.* (1873) 501–514 (1874).
7. Pennycuik, C. J. in *Avian Biology* Vol. 5 (eds Farner, D. S. & King, J. R.) 1–75 (Academic, London, 1975).
8. Rayner, J. M. V. *Curr. Ornithol.* **5**, 1–77 (1988).
9. Wieselsberger, C. Z. *Flugtech. Motorluftschiff* **5**, 225–229 (1914).
10. Butler, P. J. *et al. Comp. Biochem. Physiol. A* **126**, 379–385 (2000).

Mathematics

Where drunkards hang out

Ian Stewart

The trail of a particle undergoing brownian motion might be unkindly described as a drunken walk. A 40-year-old conjecture related to brownian motion and such random walks has finally been proved.

The drunkard's walk, more soberly known as a random walk, has long been a mainstay of probability theory. The drunkard starts from a lamppost and takes random steps forwards or backwards. Then where does he go? If the probability is the same for both directions, he will return infinitely often to the lamppost, but on average will take infinitely long to get there. Mathematicians have generalized this situation to a random walk on a regular lattice in two, three or even more dimensions. A random walk happens on a discrete lattice, but an analogous process occurs in continuous space. This is brownian motion, in which the random jiggling movements of particles in a fluid suspension are attributed to collisions with fluid molecules. A conjecture¹ related to both of these situations — first posed by Paul Erdős and S. James Taylor in 1960 — has now been proved² in a paper in *Acta Mathematica* by Amir Dembo, Yuval Peres, Jay Rosen and Ofer Zeitouni.

For random walks in higher dimensions there are some important differences. For example, the drunkard returns to the lamppost with 100% probability for walks in one and two dimensions, but with a lower probability for three dimensions or higher. If you are lost in a desert and wander at random, eventually you'll get back to where you started. But if you're 'Lost In Space', you may not. The theory of random walks goes back to the late nineteenth century and the pioneers of combinatorial probability theory. But the specific problem under discussion originated more than forty years ago, when Erdős and Taylor posed a question about random walks on a square lattice in the plane. How many times does the walker revisit the most frequently visited site in the first n steps? In other words, how many times does the drunkard go to his favourite watering hole?

Erdős and Taylor were able to make significant progress towards an answer. They

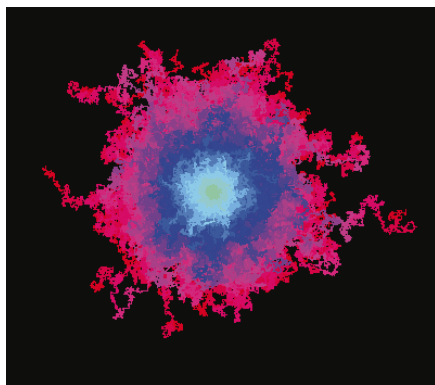


Figure 1 Random walks and fractals. A contour map of the territory covered by 500 random walkers, who started together at the centre. The different colours of the pixels reflect how often a pixel is visited. The roughness of the contour surfaces appears to be the same however many times they are visited. This 'self-similarity' is one of the defining features of fractals — the idea that if we shrink or enlarge a fractal pattern its appearance should remain unchanged. This random walk is isotropic, so its distribution has circular symmetry. The path of a particle undergoing brownian motion in the plane is analogous to a random walk, and a 40-year-old conjecture related to brownian motion and random walks has just been proved by Dembo *et al.*²

proved that for a large number of steps, n , the number of visits to the most frequently visited site lies between $(\log n)^2/4\pi$ and $(\log n)^2/\pi$. They conjectured that the larger of these numbers is the correct answer. (Curious that π should turn up here, but more on that later.) The methods used by Dembo *et al.*² to prove this conjecture are taken from fractal geometry: the core of the proof is a study of the fine multifractal structure of brownian motion. So their results constitute an important and rigorous application of fractals to probability theory and mathematical physics. Brownian motion is no longer important in its original physical context, but it has many other applications, including some to the financial sector.

In 1987, E. A. Perkins and Taylor obtained upper and lower bounds analogous to those of Erdős and Taylor for brownian motion in the plane. Given a most-frequently-visited disc of fluid, they tried to find the fraction of time during which the randomly jiggling particle is inside that disc. They proved that the answer lies between a certain expression and the same expression divided by four, and conjectured that the larger of these provides the correct answer. Their results were clear analogues of those of Erdős and Taylor. Dembo *et al.*² also prove the Perkins–Taylor conjecture, and again show that the upper bound is the correct answer. Indeed, they work mainly on brownian motion, and

then transfer their results to random walks.

These results tell us a great deal about the statistical properties of these processes. For example, mathematics can describe the probability distribution of fluctuations (departures from the average behaviour), which obey power-law statistics, and the value of the limit determined by Dembo *et al.* appears in the power laws governing their behaviour. The authors also reveal some finer detail. For example, they prove that the most frequently visited points consistently lie near the boundary of the region that the random walk visits. It seems that the drunkard's favourite hangout is as far away from the lamppost as he can get.

Last year, the same authors studied brownian motion in three dimensions, where the key insight turned out to be a localization effect. Spheres of fluid that are most frequently visited gain most of their visits during a relatively short interval of time. Essentially, the drunkard gets inside the sphere and stays there for a while. This does not happen in two dimensions — instead, the drunkard makes frequent excursions away from the most frequently occupied disc, but keeps returning to it. These excursions occur on all length scales, which is where fractal geometry comes in. The main technical issue is how the lengths of these

excursions away from the disc, on a given scale, relate to the fraction of time for which the disc is occupied. This relationship is multifractal — it can be represented by a family of fractals whose fractal dimension (a measure of their roughness) is not constant³.

Where does that π come from? As a fundamental constant in mathematics, π turns up all over the place, from geometry to number theory. Often the connection to its original definition involving circles is obscure. But in this case there is a clue in the repeated reference to occupation of discs and spheres. Brownian motion is isotropic: all directions are treated equally. So the probability distributions associated with brownian motion have circular symmetry. The same is approximately true for random walks (Fig. 1) when the number of steps, n , is large. Once circular discs enter the analysis, π cannot be far away. But this is not something the drunkard would expect to stumble upon. ■

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK. e-mail: ins@maths.warwick.ac.uk

1. Erdős, P. & Taylor, S. J. *Acta Math. Acad. Sci. Hung.* **11**, 137–162 (1960).
2. Dembo, A., Peres, Y., Rosen, J. & Zeitouni, O. *Acta Math.* **186**, 239–270 (2001).
3. Falconer, K. *Techniques in Fractal Geometry* (Wiley, Chichester 1997).

Archaeology

Tree trail to Chaco Canyon

Jared Diamond

Strontium isotopes have been used to identify the sources of timber in buildings around one thousand years old. The method can now help to solve a range of other problems.

The aim of 'provenance studies' is to identify the geographical source of a material by measuring some chemical or physical property (a 'signature') that is known to vary geographically. As they describe in *Proceedings of the National Academy of Sciences*, English *et al.*¹ have developed a method for finding the origin of plant materials by means of their strontium isotope ratio — a signature already widely used as an environmental tracer by geologists, hydrologists, ecologists and archaeologists^{2–5}. English *et al.* apply the method to an archaeological mystery: the origin of the big roof timbers at the famous ancient Native American site of Chaco Canyon (Fig. 1, overleaf), which stands in a now-treeless desert of the southwestern United States⁶. The results demonstrate not only the existence of a complex regional economic system in one part of the ancient world, but also a new method that can be applied more broadly.

Archaeologists had already worked out

other provenance methods to identify trade networks and migration routes. In particular, geographical variation in the trace-element composition and isotopic ratios of rocks and clays had been used to locate sources of stone tools and pottery. For instance, proto-Polynesian stone tools from a site on Fiji were found to be made of obsidian from the island of New Britain 4,500 km to the west, proving the existence of a long-distance trade network 3,000 years ago⁷. As for biological materials, archaeologists have attempted to identify the sources of human⁸ and animal teeth, but very little work has been done on plant materials.

The main challenge in these studies is to select a signature that varies on a geographical scale appropriate to the problem of interest. For example, discriminating between potential sources 100 km apart requires a signature that varies in the environment on that scale, rather than on a scale of 1,000 km (yielding no difference between the potential sources) or of 1 km

Figure 1 The ruins at Chaco Canyon, with (inset) doorways in one of the Anasazi great houses.



(yielding too much variation within a single source).

One of the most advanced societies in North America before the time of Columbus was the Anasazi culture centred at Chaco Canyon around a thousand years ago. The Chaco Anasazi erected a dozen buildings, termed great houses and incorporating over 200,000 logs, that were up to five storeys high and 600 rooms in extent. The area's dry climate has resulted in excellent preservation of the logs, and the year of felling can be identified exactly from their annual tree rings. For the roof beams the Anasazi required straight, tall trees that were not available nearby. So they made much use of conifers growing on mountains at least 75 km away, which yielded beams averaging 5 m long, 22 cm in diameter and weighing 275 kg. But there are three such sources of conifers nearly equidistant from Chaco Canyon: the Chuska, San Mateo and San Pedro mountains. Which did the Chaco Anasazi actually use?

As a diagnostic signature, English *et al.* selected strontium isotopes on the basis of an earlier study in another mountain range⁴. The $^{87}\text{Sr}/^{86}\text{Sr}$ ratio varies between about 0.7 and 4 with rock age and rock rubidium content, because ^{87}Sr is produced by radioactive decay of ^{87}Rb . Strontium is chemically very similar to calcium, so strontium from local atmospheric dust and from local soil produced by rock weathering is readily incorporated into plants and animals.

English *et al.* sampled 73 living trees of

six conifer species from the three mountains, as well as soil and stream water. Fortunately the three mountains proved to be clearly separated by their $^{87}\text{Sr}/^{86}\text{Sr}$ ratios, with no overlap at all, and on a single mountain the ratio did not vary significantly among tree species, stands or individuals. From six of the 12 Chaco great houses the authors then sampled 52 conifer logs, selected on the basis of their tree rings to have felling dates ranging from AD 974 to 1104, spanning at least three human generations. The result was that two-thirds of the logs could be traced to the Chuska Mountains, one-third to the San Mateo Mountains, and none at all to the San Pedro Mountains. In a given year, logs from one source could end up in several great houses: for instance, logs felled in the Chuska Mountains in 1037 were used at both Pueblo Bonito and Pueblo del Arroyo. Conversely, wood from both sources from the same year could end up in one house: room 86 at Pueblo Bonito, for instance, contains logs felled in both the Chuska and San Mateo mountains in 974. Both sources were used over the whole period from 974 to 1104.

Here we have unequivocal evidence of a well-organized, long-distance supply network for the Anasazi capital of Chaco Canyon. The synchronous use of so many beams from both sources testifies to the Chacoans' ability to mobilize large inter-community labour forces, and to coordinate the resulting flows of materials. Archaeologists had previously identified peripheral



100 YEARS AGO

A few days ago representatives of the world's science met in Berlin to do honour to one of the world's veteran men of science. The occasion of Prof. Virchow's eightieth birthday was seized by many learned societies and private individuals to express their appreciation of the great debt owed by mankind to this epoch-making thinker and worker... The work by which Virchow will always be known is his "Cellular Pathology". As Lord Lister truly remarked, workers of the present generation cannot conceive the effect which was produced upon the medical world by this book. In 1826 botanists began to regard plants as collections of cells; in fact, Schwann firmly established the position of the cell as the unit of vegetable morphology. Owing, no doubt, to the less distinctly defined characters of the animal cell, it was not until later that Kölliker and others extended the cellular theory to animal tissues. Virchow, in 1858, found a wider application for this theory and demonstrated that pathological tissues also were collections of cells, and that the phenomena of their growth was covered by the generalisation *omnis cellula a cellula*. From *Nature* 17 October 1901.

50 YEARS AGO

Very little is known about animal life in the greatest ocean depths; only two previous hauls seem to have been made below 6,000 metres. In 1899 the U.S. Fish Commission steamer *Albatross* obtained siliceous sponges in the Tonga Trench at 4,173 fm. or 7,632 metres. A similar deep haul was not made until 1948 when the Swedish Deep Sea Expedition trawled holothurians, a fragment of a polychaete, an isopod and a few amphipods in the Porto Rico Trench at depths between 7,625 and 7,900 metres. I can now report that in July the Danish Deep Sea Expedition of 1950–1952 in the research ship *Galathea* obtained living animals of several phyla from a depth of greater than 10,000 metres in the Philippine Trench, more than 2,000 metres below previous known occurrences... Prof. Claude E. ZoBell, of the Scripps Institute, University of California, has secured live samples of bacteria from the bottom deposits and is keeping cultures on board under his special high-pressure technique. The proof that a variety of animals have been adapted to life under a pressure of more than 1,000 atmospheres must lead to a renewed consideration of many problems in physiology and biochemistry. From *Nature* 20 October 1951.

Chacoan sites near the bases of the Chuska and San Mateo mountains, plus roads connecting them to the capital. It seems likely that people living at the peripheral sites cut, dried and stockpiled logs of the desired size for transport to the houses at Chaco Canyon then under construction.

English *et al.* had done their homework: the strontium isotopes had been well behaved in an earlier study on a single mountain source⁴, and they knew that strontium does not undergo biological isotopic fractionation. But they were also lucky. They had no way of knowing in advance that the spatial variation of ⁸⁷Sr/⁸⁶Sr ratio was perfect for their purpose — no variation within one mountain, but complete separation of values between mountains. These results re-emphasize the importance, in provenance studies, not only of choosing a signature with an appropriate spatial variation, but also of sampling appropriately to establish that scale of variation.

Where do we go next? Use your imagination! The most straightforward extension of the ⁸⁷Sr/⁸⁶Sr method in provenance studies will be to other ancient plant materials, and every archaeologist will have his or her own wish list. Where did the cobs of the corn come from that Chaco Canyon residents ate? Were they too imported, or were they grown in the canyon itself⁹? Is the Shroud of Turin made from fibres of plants grown in Palestine (and so possibly genuine) or in France (and so a late fake)? Whence came the

wrappings of Egyptian mummies, the roof beams of Roman buildings, and the timber of the Viking ships excavated in Norway?

There are also possible applications for strontium isotopes in forensic studies (such as detecting illegally exported elephant ivory¹⁰) and climatology (identifying secular changes in atmospheric circulation from changes in dust-deposited strontium ratios at a given site). A stickier problem that can be tackled is protecting yourself against the risk that your supposedly pure Vermont maple syrup is actually an inferior type from New Hampshire¹¹. In short, “have method, will travel” — not just geographically, but also to other uses not yet conceived. ■

Jared Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90095-1751, USA.

e-mail: jdiamond@mednet.ucla.edu.

1. English, N. B., Betancourt, J. L., Dean, J. S. & Quade, J. *Proc. Natl Acad. Sci. USA* **98**, 11891–11896 (2001).
2. Quade, J., Roe, L., DeCelles, P. G. & Ojha, T. *Science* **276**, 1828–1831 (1997).
3. Bullen, T. D., Krabbenhoft, D. P. & Kendall, C. *Geochim. Cosmochim. Acta* **60**, 1807–1821 (1996).
4. Graustein, W. C. & Armstrong, R. L. *Science* **219**, 289–292 (1983).
5. Price, T. D., Manzanilla, L. & Middleton, W. D. *J. Archaeol. Sci.* **27**, 903–913 (2000).
6. Vivian, R. G. *The Chacoan Prehistory of the San Juan Basin* (Academic, New York, 1990).
7. Best, S. *Archaeol. Oceania* **22**, 31–32 (1987).
8. Price, T. D., Johnson, C. M. & Ezzo, J. A. *J. Archaeol. Sci.* **21**, 315–330 (1994).
9. Cordell, L. S., Durand, S. R., Antweiler, R. C. & Taylor, H. E. *J. Archaeol. Sci.* **28**, 501–513 (2001).
10. Van der Merwe, N. J. *et al. Nature* **346**, 744–746 (1990).
11. <http://www.dartmouth.edu/~dreeder/fraud.html>

material whose magnetism is based on the properties of its component molecules rather than its atoms — was reported in 1972. This was followed in 1986 by the discovery of ferromagnetism involving *p*-electrons in an iron-containing organic-based material⁴. This was an important step forward because magnetism in metal-free compounds must involve electrons from *p*-atomic orbitals. Magnetic ordering in systems solely based on *p*-electrons was considered impossible 20 years ago. The first metal-free organic magnet was discovered ten years ago⁵. It was a derivative of nitronyl nitroxide that became magnetically ordered at 0.65 K, a very low *T_C*.

Buckminsterfullerene⁶, or fullerene for short, is an allotropic form of carbon, like graphite or diamond. It consists of 60 carbon atoms forming a quasi-spherical molecule, which under normal conditions has no intrinsic magnetic moment. For a magnetic moment to exist, an electron must be transferred to C₆₀ from a donor molecule. At the same time as the discovery of the first organic magnet, ferromagnetism was also reported in a fullerene-based charge-transfer salt⁷, TDAE·C₆₀, which has a Curie temperature of 16 K. A second example⁸ of a fullerene-based ferromagnet was discovered in a cobaltocene-doped fullerene derivative, which has a *T_C* of 19 K. In this case, only the fullerene molecules have magnetic spins, and so are solely responsible for the observed magnetism. Orientational order is also crucial for achieving ferromagnetism in fullerene derivatives⁹.

Despite their best efforts, scientists have so far failed to observe magnetic ordering at room temperature and above in a metal-free organic compound. The highest ordering temperature reported to date for an organic magnet is 36 K for a sulphur-based free radical¹⁰. Under a pressure of 16 kilobars¹¹ this temperature can be raised to 65 K — still far below room temperature of 300 K. For years there have been many reports of observations of weak spontaneous magnetization above room temperature in heat-treated organic compounds. In virtually every case, it has been difficult to establish the intrinsic character, or even the reproducibility, of the observed magnetism¹². The possibility that some interesting magnetic behaviour might be going on behind those observations has never been addressed systematically.

This is precisely what Makarova *et al.*¹ have now done. Following their observation of unexpected magnetic behaviour in a C₆₀ sample, which had been polymerized at high temperatures and pressures, they concentrated on its magnetic properties. Up to every reasonable limit they exclude the possibility of impurities being the origin of the observed magnetism, and they also demonstrate the reproducibility of the effect. High-pressure and high-temperature treatment

Materials science

A magnet made from carbon

Fernando Palacio

Conventional wisdom says that magnetic materials have to contain some metallic atoms. So the discovery of a type of pure carbon that is magnetic at room temperature is bound to invite controversy.

Magnets have played an important role in civilization since the earliest times. Vikings, for example, navigated their ships using a fish-shaped piece of lodestone on a floating support as a compass. Today, from electricity generation to small kitchen appliances, not to mention cars, headphones and laptop computers, magnets are essential components of our everyday life. Traditional materials used for making magnets include iron and iron oxides (lodestone and ferrites).

But scientists are also interested in developing magnets from molecular materials whose constituent atoms are non-metallic. Such metal-free magnets would be electrical insulators (reducing energy losses in some applications) and should be cheaper and lighter than their metallic counterparts. Although a handful of metal-free magnets have been discovered to date, their magnetic

properties occur only at very low temperatures. Now on page 716 of this issue Makarova *et al.*¹ describe the discovery of spontaneous magnetization in a pure carbon material of polymeric C₆₀, which is magnetic at room temperatures and higher (up to 500 K).

All magnets lose their magnetic character above some critical temperature, *T_C* for short, which in the case of ferromagnets is also known as the Curie temperature. Below *T_C* the magnetic moments, or ‘spins’, of electrons in the material order in such a way that a net magnetization arises. For some materials, namely ferromagnets such as iron, magnetic ordering aligns all the moments parallel to each other. But this is only the simplest magnetic structure for a magnet, and many other more complex structures are also possible.

The first molecular ferromagnet^{2,3} — a

can transform C_{60} from a crystalline state of isolated molecules, held together only by weak van der Waals forces, into polymeric phases in which the molecules are covalently bonded to each other. One-, two- and three-dimensional polymers can be formed by modifying the treatment conditions.

In this study, Makarova *et al.*¹ prepare a two-dimensional rhombohedral phase, which resembles highly oriented heat-treated graphite, but with layers of covalently bonded C_{60} molecules instead of graphite. It is only in this rhombohedral phase that a very weak magnetization — they are probably too optimistic in calling it ferromagnetism — is observed. The appearance of magnetic behaviour seems to be very sensitive to the pressure and temperature conditions under which the samples are prepared. Spontaneous magnetization can be made to disappear by heating the sample at 700 K for several hours. It seems, therefore, that the observed magnetism is intrinsic to the samples, although instinct might tempt one to think otherwise.

Experimental facts apart, I anticipate that the results reported by Makarova *et al.*¹ have the right mix of ingredients to be controversial. There are important questions which still remain open and that will require proper attention. The most basic one is also the most obvious: where are the magnetic moments? The authors speculate that the moments

might arise from defects caused by broken fullerene bonds or, alternatively, that some covalent bonds require only one electron; the second electron would then be responsible for the magnetic ordering. A second question concerns the magnetic ordering itself. The spontaneous magnetization is very small, about a hundred times smaller than the magnetization expected from the ferromagnetic alignment of one magnetic moment per molecule of C_{60} .

If confirmed, this result will represent a breakthrough in the magnetism of metal-free materials. We can expect, and should welcome, a lively discussion, which will help stimulate research into understanding the behaviour of this unusual carbon magnet. ■

Fernando Palacio is in the Instituto de Ciencia de Materiales de Aragón, CSIC - Universidad de Zaragoza, 50009 Zaragoza, Spain.
e-mail: palacio@posta.unizar.es

1. Makarova, T. L. *et al.* *Nature* 413, 716–718 (2001).
2. Wickmann, H. H. *J. Chem. Phys.* 56, 976–982 (1972).
3. DeFotis, G. C., Palacio, F., O'Connors, C. J., Bhatia, S. N. & Carlini, R. L. *J. Am. Chem. Soc.* 99, 8314–8315 (1977).
4. Miller, J. S. *et al.* *J. Chem. Soc. Chem. Commun.* 1026–1028 (1986).
5. Tamura, M. *et al.* *Chem. Phys. Lett.* 186, 401–404 (1991).
6. Allemand, P. M. *et al.* *Science* 253, 301–303 (1991).
7. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* 318, 162–163 (1985).
8. Mizel, A. *et al.* *Chem. Phys. Lett.* 298, 329–334 (1998).
9. Narymbetov, B. *et al.* *Nature* 407, 883–885 (2000).
10. Palacio, F. *et al.* *Phys. Rev. Lett.* 79, 2336–2339 (1997).
11. Mito, M. *et al.* *Polyhedron* 20, 1509–1512 (2001).
12. Miller, J. S. *Adv. Mater.* 4, 435–438 (1992).

Huntington's disease

Exploiting expression

Gillian P. Bates

Huntington's disease results from defects in the huntingtin protein, but the exact mechanism has been unclear. Researchers now have a better idea, and the knowledge has proved beneficial — for flies at least.

Close to the start of the huntingtin gene lies a repetitive sequence of nucleotides, CAGCAGCAG..., that is potentially deadly. This sequence codes for a stretch of glutamine amino acids at one end of the huntingtin protein. People with more than 40 glutamines will develop Huntington's disease; those with fewer than 38 will be unaffected. On page 739 of this issue, Steffan and colleagues¹ shed light on how the elongated polyglutamine tract leads to the progressive degeneration and death of neurons that characterizes this disorder. Their results show that the expanded tract interferes with the apparatus for switching on and regulating genes, and so perturbs gene expression. Moreover, the authors use both genetic and pharmacological approaches to compensate for such interference, preventing neurodegeneration and premature death in a fruitfly model of Huntington's disease.

At the last count there were nine neuro-

degenerative diseases that are caused by the expansion of a polyglutamine tract in one protein or another. All of these disorders are associated with the formation in nerve cells of insoluble aggregates containing the affected protein. Stretches of polyglutamine do not always become harmful, however — they also occur normally in many of the proteins that make up transcription-factor complexes, which regulate gene expression. For example, an acetyltransferase enzyme known as CBP has a tract of 18 glutamines. Acetyltransferases activate transcription by adding acetyl groups to histones — proteins that help to package DNA into a compact form. In organisms except bacteria, transcription is regulated by the interplay between histone acetyltransferases and histone deacetylases (Fig. 1a, overleaf).

Two years ago it was reported that CBP can be recruited into aggregates of polyglutamine-containing proteins in mammalian

cells, and that this depends on the presence of CBP's polyglutamine tract². Since then the enzyme has been identified as a component of intranuclear aggregates in mouse models of Huntington's disease and in the autopsied brains of patients who suffered from the disorder^{3,4}. Together with the fact that replenishing CBP reverses toxicity in mammalian cell models of polyglutamine diseases^{4,5}, these results raised the possibility that the recruitment of CBP by polyglutamine-containing proteins contributes to the symptoms of these disorders. Steffan *et al.*¹ provide further evidence of this, at least for Huntington's disease — but an additional mechanism is also at work.

The authors previously found³ that the amino terminus of huntingtin (the part that contains the polyglutamine tract) can interact with CBP without forming large aggregates. They now show¹ that an amino-terminal fragment of huntingtin binds directly to CBP and another protein, P/CAF (p300/CBP-associated factor) — but through their acetyltransferase domains, not their polyglutamine tracts. This results in a reduction in the acetyltransferase activity of CBP, p300 and P/CAF in cell-free assays, and a decrease in the level of histone acetylation in mammalian cells. In both cases, histone acetylation is reduced irrespective of the length of the polyglutamine tract, but becomes more pronounced with longer tracts. A reduction in acetylation induced by lengthened polyglutamine tracts has also been seen in yeast and other mammalian cells (B. Hughes and A. McCampbell, personal communication).

Steffan *et al.*¹ propose that expression of the amino-terminal fragment of huntingtin leads to a decrease in histone acetylation that would result in a decrease in gene transcription (Fig. 1b). This hypothesis is consistent with the downregulation of specific genes in a mouse model of Huntington's disease⁶. The authors also reasoned that the effects of the fragment might be redressed, at least in part, by using compounds that inhibit the cellular histone deacetylases — the reduction in deacetylase activity should compensate for the loss of acetyltransferase activity (Fig. 1c).

To test this idea, Steffan *et al.* first used several pharmacological inhibitors of histone deacetylases, and found that they do indeed increase acetylation in mammalian cells in culture. They then turned their attention to mutant flies that expressed in their neurons either a huntingtin fragment with 93 glutamines or a peptide consisting simply of 48 glutamines. In adult flies the huntingtin fragment causes neurodegeneration, and the peptide leads to early death. When the flies were reared on food containing histone-deacetylase inhibitors, the level of neurodegeneration was much reduced and the flies lived for longer. Even when the animals already showed signs of

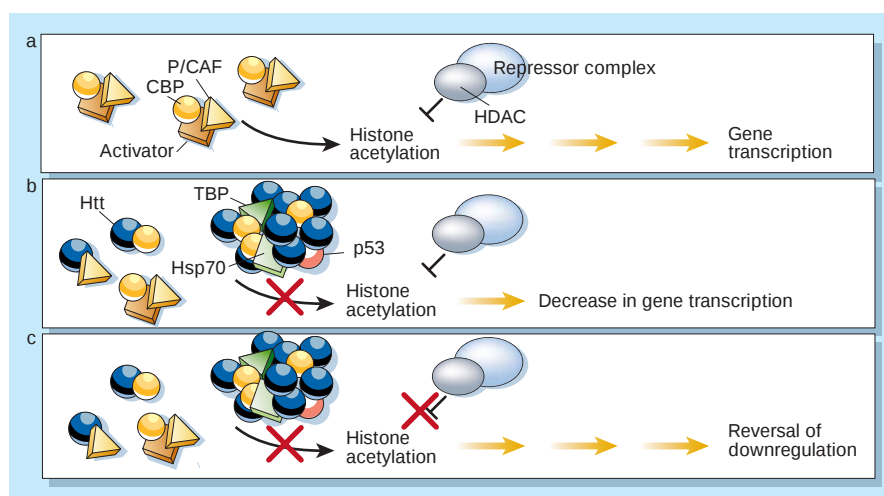


Figure 1 Huntington's disease and gene expression. **a**, Gene transcription is switched on by complexes that contain activator proteins and acetyltransferases (such as CBP and P/CAF); acetyltransferases add acetyl groups to the histone proteins that package DNA into a compact form. Transcription is regulated by an interplay between histone acetyltransferases and histone deacetylases (HDACs). **b**, Steffan *et al.*¹ have found that mutant huntingtin protein (Htt) inhibits histone acetylation by blocking histone acetyltransferases or recruiting them into aggregates. The reduction in histone acetylation leads to decreased transcription. **c**, The balance can be partially redressed by using molecules that prevent histone deacetylation.

neurodegeneration, the inhibitors dramatically slowed or stopped further degeneration. Similar results were obtained when the level of histone-deacetylase activity was reduced by a genetic approach. So the authors have shown — at least in invertebrates — that reducing the level of histone acetylation has a major role in the pathogenesis of Huntington's disease.

It seems strange that the impairment of transcription factors expressed from early

development throughout an embryo can cause adult-onset diseases that affect only neurons. But concerns about the validity of the findings should be laid to rest by the discovery that a polyglutamine expansion in a key transcription factor, TBP, directly causes a nerve-cell-specific disease, spinocerebellar ataxia 17 (ref. 7). Furthermore, cells are extremely sensitive to levels of CBP, as witnessed by the fact that Rubinstein–Taybi syndrome, a rare developmental disorder, is

caused by mutations, that decrease histone acetylation, in just one of the two CBP genes⁸.

It is clear that huntingtin binds to acetyltransferases such as CBP, but it is less certain how it reduces acetyltransferase activity (Fig. 1b). One answer could be that acetyltransferases are simply recruited into aggregates of huntingtin and so are prevented from binding to genes. This is an attractive idea because the propensity of huntingtin to form

Oceanography

Quo vadis, iceberg?

Ever since the sinking of the *Titanic*, icebergs have aroused fear and fascination among seafarers as well as the public. They apply an enormous force to anything that stands in their way. But they also transport immense amounts of fresh water, inspiring the suggestion that water could be delivered to dry regions of the world in the form of icebergs.

To understand the paths of individual icebergs around Antarctica, Rupert M. Gladstone of the University of East Anglia, UK, and his colleagues have built a numerical iceberg-trajectory model. In the *Journal of Geophysical Research* (**106**, 19903–19915; 2001), they report that only in the Ross Sea, the Weddell Sea and at the Kerguelen Plateau are the trajectories likely to leave the coast — in line with observations.

The researchers fed climatological data of iceberg calving fluxes at 29 points along the coast of Antarctica into their model. Wind and water drag the simulated icebergs (assumed to measure

between 60 and 2,200 m across and about 250 m deep) anticlockwise along the coast, while waves erode the icebergs at the water surface and turbulent heat flux melts the underwater part.

A calculation of the total injection of fresh water from melting icebergs indicates an influence on the same order of magnitude as the difference between precipitation and evaporation along the Antarctic coast and in some open ocean regions, particularly off the Antarctic Peninsula.

Melting icebergs cool and freshen the surface ocean. On balance, this process produces a more stable water column, less likely to allow the formation of deep water that drives the global ocean circulation.

But before jumping to conclusions about the importance of their model's results for global climate, the researchers hasten to warn us that they omitted to model 'giant icebergs', which are larger than 18.5 km across and may



account for as much as half the total Antarctic iceberg calving flux. Because of their sheer size, these giants have quite different dynamics to normal icebergs, and

it is not clear whether they are more likely to remain close to the coast or to reach the open ocean. So maybe we have just seen the tip of the iceberg.

Heike Langenberg

TPL

aggregates correlates with the length of its glutamine tracts⁹, and the onset, symptoms and pathology of disease would then be related to the appearance and distribution of aggregates. But Steffan *et al.* also found that the amino-terminal part of huntingtin directly inhibits several acetyltransferases in the absence of large aggregates. Here, the rate-limiting step in the onset and progression of symptoms might be, for example, the movement of huntingtin to the nucleus.

Whatever the mechanism, the use of histone-deacetylase inhibitors had remarkable effects in an invertebrate model of Huntington's disease¹. Several such inhibitors are already in clinical use, or are in early clinical trials, for treating patients with the disorder¹⁰. But we need as much information as possible about which inhibitors are likely to

be most effective, so it should also be a high priority to test them in mouse models. ■

Gillian P. Bates is in the Division of Medical and Molecular Genetics, GKT School of Medicine, King's College, 8th Floor Guy's Tower, Guy's Hospital, London SE1 9RT, UK.

e-mail: gillian.bates@kcl.ac.uk

1. Steffan, J. S. *et al.* *Nature* **413**, 739–743 (2001).
2. Kazantsev, A., Preisinger, E., Dranovsky, A., Goldgaber, D. & Housman, D. *Proc. Natl Acad. Sci. USA* **96**, 11404–11409 (1999).
3. Steffan, J. S. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 6763–6768 (2000).
4. Nucifora, F. C. Jr *et al.* *Science* **291**, 2423–2428 (2001).
5. McCampbell, A. *et al.* *Hum. Mol. Genet.* **9**, 2197–2202 (2000).
6. Luthi-Carter, R. *et al.* *Hum. Mol. Genet.* **9**, 1259–1271 (2000).
7. Nakamura, K. *et al.* *Hum. Mol. Genet.* **10**, 1441–1448 (2001).
8. Murata, T. *et al.* *Hum. Mol. Genet.* **10**, 1071–1076 (2001).
9. Scherzinger, E. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 4604–4609 (1999).
10. Marks, P. A., Rifkind, R. A., Richon, V. M. & Breslow, R. *Clin. Cancer Res.* **7**, 759–760 (2001).

Crystallization

Diversity suppresses growth

David W. Oxtoby

Colloids, which consist of small particles in suspension, can switch from a fluid to a crystalline state. But a careful simulation of this phase transition shows that some types of colloids cannot crystallize.

The shimmering iridescence of an opal gemstone arises from the diffraction of visible light by tiny particles. Such opalescence can occur because colloidal particles, which are hundreds to thousands of times larger than single atoms, can spontaneously arrange themselves into regular crystalline arrays — just like atomic or molecular crystals. In the laboratory, such

colloidal crystals form even when the particles repel each other. Theorists love colloidal crystals because they can imitate an idealized crystal structure formed from nearly perfect hard spheres; experimentalists love them because the particles are large enough, and the growth process slow enough, for crystallization to be observed in detail.

On page 711 of this issue, Auer and

Frenkel¹ describe some surprising results from a simulation of colloids that challenges classical theories of crystal formation. They predict that if the range of sizes (polydispersity) of the hard colloidal particles in suspension is large enough, the mixture will be truly amorphous, and so will be unable to form even the smallest crystals. This result has potentially important consequences for the design of new materials.

The earliest stage of crystallization — referred to as nucleation — is the hardest to observe, yet it usually dictates the rest of the growth process. Technically, nucleation is the initial appearance of a new phase (here, the emergence of tiny crystal nuclei) during a first-order phase transition, such as boiling or melting. There is always a 'free-energy barrier' to nucleation, because creating a new interface between different phases costs free energy — a thermodynamic quantity that reflects the amount of energy available for work when the system undergoes a change. The free-energy costs are higher for small nuclei because they have a higher ratio of surface area to bulk volume, so small crystallites have to reach a certain size — the 'critical nucleus' — to stand a chance of surviving.

Despite the free energy required to form the critical nucleus, once it appears its subsequent growth is thermodynamically favoured. According to the classical theory of nucleation², the height of the nucleation free-energy barrier ΔG^* is proportional to $\gamma^3/|\Delta\mu|^2$, where γ is the free energy of the solid–liquid interface and $\Delta\mu$ is the difference in chemical potential between the crystal and fluid. Because $\Delta\mu$ increases with density (from its zero value at phase



Figure 1 Colloidal crystals. A series of hard-sphere colloids shows how crystal structure can change with increasing particle density⁸. The most dilute sample (extreme right) shows no crystallization, whereas samples with higher particle concentrations crystallize to varying degrees. Colloids at intermediate concentrations form small crystallites, which settle under gravity to create a boundary between the colloidal fluid and crystal phases. In three samples the entire

volume is filled with crystallites, which are either homogeneous (small regular crystallites) or heterogeneous (large irregular crystals). The most concentrated samples (extreme left) remained in a largely amorphous state for several months. In a new computer simulation, Auer and Frenkel¹ show that a range of colloidal particle sizes (polydispersity) can also hinder crystallization. (Image kindly provided by P. N. Pusey.)

equilibrium), the natural assumption is that the barrier to nucleation must steadily decrease. But Auer and Frenkel have demonstrated¹, using advanced computer simulation techniques³, that the free-energy barrier preventing the formation of crystals can increase, even though the thermodynamic driving force favouring the crystalline phase continues to increase. So, at high enough densities, ΔG^* passes through a minimum and begins to increase, a result not predicted by classical theory. This effect is strongest for polydisperse systems.

An increase in the barrier to nucleation implies that there is a decrease in the rate of nucleation. This in itself is not surprising, because in experiments the crystal nucleation rate for many materials goes through a maximum and then decreases at high densities or low temperatures (Fig. 1). This effect is usually attributed to a dramatic slowdown in the kinetics, because fluid motion becomes sluggish as the density increases or the temperature decreases⁴. Auer and Frenkel's work shows that this interpretation may not always be correct — the maximum in the crystal nucleation rate may come from a minimum in the free-energy barrier preventing formation of the critical nucleus. This prediction represents yet another challenge⁵ to classical nucleation theory.

There are two further intriguing and puzzling aspects of the work by Auer and Frenkel¹. The first is suggested by an application of the nucleation theorem to their data. The nucleation theorem is a general result from the thermodynamics of small systems that states that the rate at which the nucleation free-energy barrier changes with chemical potential can be used to predict the excess number of particles in the critical nucleus relative to the background fluid⁶ — in other words, how dense the critical nucleus is. According to Auer and Frenkel, for ΔG^* to pass through a minimum the excess number of particles must pass through zero and become negative, so the density of the critical crystalline nucleus must become lower than that of the fluid around it. This prediction is surprising, because hardly any fluids have crystals that are less dense than their liquid states (water is a rare exception — ice floats). This new idea can be tested by experiment and by further simulation.

A second surprise is the suggestion that the minimum nucleation energy barrier may apply for systems of monodisperse hard spheres, as well as polydisperse ones. As a result, the density of the critical nucleus of the hard-sphere crystal would be lower than that of the surrounding liquid. Systems of monodisperse hard spheres have been studied so thoroughly that theorists thought they understood them completely, but this idea has never been proposed before. Calculations based on density functional theory⁷ —

a method better suited to the low nucleation barriers in the monodisperse case — may shed light on this question.

Auer and Frenkel also predict that a minimum nucleation rate will affect the typical size of the microcrystals that eventually result once the phase transition is complete. Although the nucleation step is difficult to observe in the laboratory, the final distribution of crystal sizes is relatively easy to measure. Moreover, it has a significant effect on the properties of the resulting solid materials. If real colloidal systems (beyond the hard-sphere model considered here) can be made fully amorphous by having sufficiently polydisperse particles, new types of materials may be created. For example, an increase in the polydispersity of synthetic semiconductor colloids might lead to amorphous

materials with new electronic properties. Studies of polydisperse colloids may also give us a better understanding of how metal alloys crystallize. There's more to colloids than meets the eye.

David W. Oxtoby is in the James Franck Institute and the Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, USA.

e-mail: d-oxtooby@uchicago.edu

1. Auer, S. & Frenkel, D. *Nature* **413**, 711–713 (2001).
2. Turnbull, D. & Fisher, J. C. *J. Chem. Phys.* **17**, 71–73 (1949).
3. Van Duijnveldt, J. S. & Frenkel, D. *J. Chem. Phys.* **96**, 4655–4668 (1992).
4. MacFarlane, D. R., Kadiyala, R. K. & Angell, C. A. *J. Phys. Chem.* **87**, 1094–1095 (1983).
5. Yau, S.-T. & Vekilov, P. G. *Nature* **406**, 494–497 (2000).
6. Oxtoby, D. W. & Kashchiev, D. *J. Chem. Phys.* **100**, 7665–7671 (1994).
7. Oxtoby, D. W. *Nature* **347**, 725–730 (1990).
8. Pusey, P. N. & van Megen, W. *Nature* **320**, 340–342 (1986).

Molecular biology

RNA enzymes for RNA splicing

Andy Newman

Many messenger RNAs are not functional until they are processed by a complex called the spliceosome. It seems increasingly likely that processing is catalysed by the RNA — and not the protein — parts of this complex.

The discovery of catalytic RNAs in 1982 forced proteins to relinquish their status as the only molecules that can catalyse chemical reactions in living cells. For instance, RNA forms the active site in the ribosome, the complex molecular machine that catalyses protein synthesis. And there are hints that an RNA-processing reaction — the splicing of precursor mRNA molecules — might be catalysed by the RNA components of the 'spliceosome' (reviewed in ref. 1). But without precise structural information about

the spliceosome's active site this has been hard to confirm. On page 701 of this issue, Valadkhan and Manley² provide further support for the idea, with their discovery that a protein-free complex of just two of the spliceosome's RNA components catalyses a reaction that is closely related to the first catalytic step of splicing.

So what is this splicing reaction? Most nuclear genes have a discontinuous structure, consisting of protein-coding sequences (exons) that are interrupted by non-coding

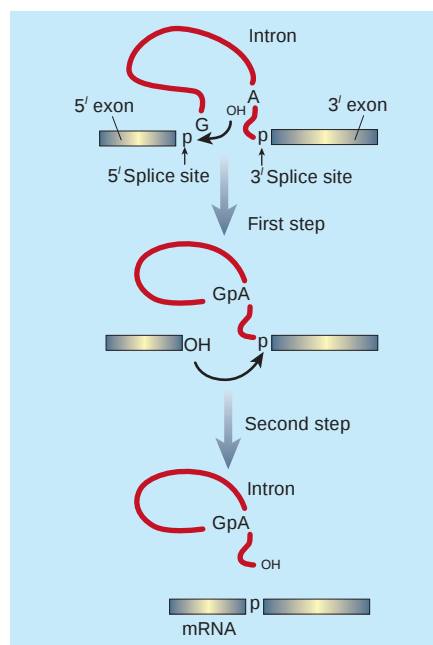


Figure 1 **Splicing of precursor messenger RNA molecules.** In the first step, the 2' hydroxyl group (OH) of an adenosine nucleotide in the intron (the branch-point adenosine, A) attacks the phosphodiester bond (p) at the 5' splice site. This bond breaks, and a new phosphodiester bond is formed (GpA). In the second step, the 3' OH group of the free 5' exon attacks the 3' splice site. The result is joined exons (the spliced messenger RNA) and a free intron.

sequences (introns). Transcription of these genes produces precursor mRNAs (pre-mRNAs), which must be processed by accurately removing the introns and splicing together the exons. These reactions are carried out by spliceosomes, which are built from five RNA molecules (known as small nuclear RNAs, snRNAs) and more than 50 proteins.

The two catalytic steps of splicing are both 'phosphotransfer' reactions — in each step, one phosphodiester bond in the pre-mRNA substrate is broken and another is formed. Picture a pre-mRNA molecule that consists simply of an exon, followed by an intron, and then another exon (Fig. 1). The two ends of an RNA are referred to as 5' and 3'; so, in our simple pre-mRNA molecule, the 5' splice site is the end of the intron that is connected to the 5' exon. In the first splicing reaction, the 5' splice site is attacked by a 2' hydroxyl group from an adenosine nucleotide that lies in a specific sequence — the 'branch-point' — in the intron. This results in the formation of a phosphodiester linkage within the intron, and a hydroxyl group at the newly exposed end of the 5' exon. In the second step, this 'new' hydroxyl group targets the intron's 3' splice site to yield the final products: spliced mRNA and the excised intron.

Of the five snRNA molecules in the spliceosome, U2 and U6 have emerged as the prime candidates for components of the spliceosome's active site. In U6, several nucleotides are crucial for the catalytic steps of pre-mRNA splicing; these nucleotides are identical in U6 molecules from a wide range of organisms¹. They include an 'AGC' triad, which is also an essential element of the catalytic domain of many self-splicing RNAs. In catalytically active spliceosomes, U2 and U6 base pair together and form a network of interactions with the branch-point and 5'-splice-site sequences^{1,3,4}.

Last year, Valadkhan and Manley⁵ showed that human U2 and U6 snRNAs anneal together, in the absence of proteins and the presence of magnesium ions, to form a stable, base-paired structure. Crucially, ultraviolet irradiation of this complex produced a specific covalent crosslink between U2 and U6, pointing to a three-dimensional structure similar to that thought to exist in catalytically active spliceosomes. These authors² have now found that this U2–U6 RNA structure forms a specific complex with a short artificial RNA molecule that contains a branch-point sequence, mimicking the helix formed between U2 and the branch-point sequence in an intact spliceosome. Moreover, this model complex activates the 2' hydroxyl of the branch-point adenosine to attack a specific phosphodiester bond in the essential AGC triad in U6 snRNA.

Of course in the spliceosome itself, the target of this attack would be the substrate's

5' splice site, which is not present in this model complex. Nevertheless, this unusual reaction has clear similarities to the first catalytic step in authentic spliceosomes. It requires specific, invariant sequences in U6, and the attacking 2' hydroxyl group is provided by an adenosine that has been 'bulged' out of a base-paired helix formed between U2 snRNA and the branch-point RNA. Interestingly, there is a precedent for the branch-point 2' hydroxyl attacking U6 snRNA rather than the 5' splice site: this is what happens in nematode spliceosomes when U6 contains specific mutations believed to cause distortion of the catalytic site⁶.

The idea that a U2–U6 complex could form the active site for the first step of splicing dovetails neatly with earlier findings that the spliceosome is metal-ion-dependent, like several other RNA enzymes that use metal ions as essential cofactors⁷. In fact, it has been shown that a specific phosphodiester linkage in U6 snRNA binds a crucial magnesium ion, which may participate directly in the chemistry of the first catalytic step of splicing⁸.

Are we to conclude that all of the spliceosome's protein components are relegated to peripheral roles away from the chemistry of the splicing reactions? Not necessarily. It remains to be seen whether the catalytic repertoire of spliceosomal RNAs extends to other chemical reactions that could be related to the second step of splicing. In any case, there is compelling evidence that at least one highly conserved spliceosomal protein, Prp8, is intimately associated with the snRNAs and the substrate in the spliceosome's active site. Overall, the current evidence suggests that Prp8 may act as a cofactor, perhaps stabilizing the RNA structures required for catalysis, rather than contributing directly to catalysis (reviewed in ref. 9).

Eventually, detailed structural information about the spliceosome's active site may help to settle the issue, as it has for the ribosome¹⁰. In the meantime, model systems such as those described here² could be the fast track to learning more about how the spliceosome does what it does. ■

Andy Newman is in the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

e-mail: newman@mrc-lmb.cam.ac.uk

1. Nilsen, T. W. in *RNA Structure and Function* (eds Simons, R. W. & Grunberg-Manago, M.) 279–307 (Cold Spring Harbor Laboratory Press, New York, 1998).
2. Valadkhan, S. & Manley, J. L. *Nature* **413**, 701–707 (2001).
3. Madhani, H. D. & Guthrie, C. *Genes Dev.* **8**, 1071–1086 (1994).
4. Sun, J. S. & Manley, J. L. *Genes Dev.* **9**, 843–854 (1995).
5. Valadkhan, S. & Manley, J. L. *RNA* **6**, 206–219 (2000).
6. Yu, Y.-T., Maroney, P. A. & Nilsen, T. W. *Cell* **75**, 1049–1059 (1993).
7. Sontheimer, E. J., Sun, S. & Piccirilli, J. A. *Nature* **388**, 801–805 (1997).
8. Yean, S.-L., Wuenschell, G., Termini, J. & Lin, R.-J. *Nature* **408**, 881–884 (2000).
9. Collins, C. A. & Guthrie, C. *Nature Struct. Biol.* **7**, 850–854 (2000).
10. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. *Science* **289**, 920–930 (2000).

Daedalus

A word in your ear

Aurorae and meteorites are usually silent, but occasionally they make a noise.

Daedalus reckons that this is because they perturb the magnetosphere, which sometimes forms an electromagnetic wave that converges on an observer. That observer is then exposed to a fluctuating magnetic field. The oxygen of the local air (which is paramagnetic) varies in volume in sympathy with the field, making the noise. The effect propagates at the speed of microwaves, not that of sound, so even a distant aurora or meteorite makes a noise while it is still visible.

So, says Daedalus, here is a method of creating sound at a distance. A parabolic aerial, or several of them in a phased array reinforcing at a chosen point, could launch a microwave beam whose local intensity would vary at an audio frequency. Its magnetic component would then affect the local air, and make a noise. The microwave frequency itself, far too high to hear, would not matter. Like an amplitude-modulated signal, the audio effect would depend only on its changing absolute intensity. Here is a way of generating truly local sound.

The obvious use is crowd control. Many popular events these days, such as pop festivals, take place in large fields or open spaces ringed by security staff. With a proper array of peripheral aerials, just one chosen spot could be irradiated with a sound signal, controlled by the security men on their watchtowers. Local revellers acting suspiciously could be warned off without any unsettling public declaration; only those being addressed would hear anything. Another use would be in supermarkets, museums and so on, where potential customers near a product or exhibit could be told all about it by a private local voice. If they drifted away, this would rapidly fade.

Indeed, says Daedalus, many public-address tasks could be done better by his microwave system. A single railway platform or section of platform could be addressed by an announcer. People using public telephones in airports, currently deafened by public-address announcements, could be spared by shaping the public microwave pattern to avoid the phone boxes. Even in a domestic setting, the hi-fi enthusiast might enjoy a purely local signal. His mobile phone might be made blessedly local too, although sadly his replies, as genuine sound, would still fill the room. For even Daedalus can think of no way of switching off a real local sound source.

David Jones

Energy saving in flight formation

Pelicans flying in a 'V' can glide for extended periods using the other birds' air streams.

Many species of large bird fly together in formation, perhaps because flight power demands and energy expenditure can be reduced when the birds fly at an optimal spacing^{1–3}, or because orientation is improved by communication within groups⁴. We have measured heart rates as an estimate of energy expenditure in imprinted great white pelicans (*Pelecanus onocrotalus*) trained to fly in 'V' formation, and show that these birds save a significant amount of energy by flying in formation. This advantage is probably a principal reason for the evolution of flight formation in large birds that migrate in groups.

We trained eight great white pelicans to fly after a moving motor boat and an ultralight aeroplane in Djoudj National Park, Senegal (Fig. 1). All flight sessions were filmed with a digital camera to measure wing-beat frequency and to synchronize recordings of behaviour and heart rate. The average heart rate was recorded by using an

electronic heart-rate logger (Polar Electro, Finland) adapted for large birds⁵; the unit was attached to the back feathers with adhesive tape and the electrodes were placed dorsally about 250 mm apart directly under the skin. Heart rates and wing-beat frequencies recorded during the different types of flight were averaged for each individual over one to four different bouts that lasted for 30 seconds to 2 minutes.

Pelicans flying over water in diagonal formation used the typical flight pattern of pelicans in formation — a few wing-beats, followed by a short, 1–2-second glide — in which 42.2% of the time is spent flapping. Birds flap in time with the leader, either in unison or in regular succession. Within the formation, the number of glides per minute was similar for all birds (average, 12.6 ± 1.6 glides per minute), but the average wing-beat frequency decreased from the leader to the bird in the fourth position (Fig. 2).

When flying alone, 1 m above water or at higher altitude, pelicans beat their wings more frequently than birds flying in formation (Wilcoxon paired signed-rank test: $Z = 2.2$, $n = 6$ birds, $P = 0.028$; and $Z = 2.0$, $n = 5$, $P = 0.043$, respectively), but still glided briefly. When in formation, birds had a heart rate that was 11.4–14.5% lower than in birds flying alone at 1-m or 50-m altitude (Fig. 2; $Z = 2.0$, $n = 5$, $P = 0.043$; and $Z = 2.4$, $n = 6$, $P = 0.018$, respectively). When gliding, birds' heart rates decreased further and they beat their wings only occasionally (Fig. 2; $Z = 1.8$, $n = 7$, $P = 0.068$ for both heart rate and wing-beat frequency). Birds flying alone over water had a slightly higher mean wing-beat frequency and lower heart rates than birds flying at higher altitude (Fig. 2); however, the differences were not significant ($Z = 1.1$, $n = 6$, $P = 0.249$; and $Z = 0.1$, $n = 5$, $P = 0.893$, respectively), suggesting that any ground effect⁶ was not significant.

Our results provide empirical evidence that, compared with solo flight, formation flight confers a significant aerodynamic advantage which allows birds to reduce their energy expenditure while flying at a similar speed. In birds flying in formation, each wing moves in an upwash field that is generated by the wings of the other birds in the formation. Modelling has shown that when birds are flying with optimal spacing, a maximal reduction in power can be achieved³ and total transport costs can be substantially reduced². However, field observations of V formations indicate that birds often shift from their optimal positioning, perhaps in an attempt to maximize



Figure 1 Great white pelicans flying in formation over a river (photograph, R. Marzin, Galatée Films).

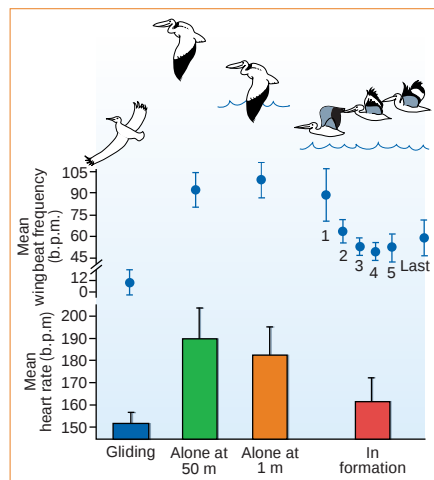


Figure 2 Wing-beat frequency and heart rate of pelicans engaged in various types of flight (mean $\pm$ 1 s.d.). Birds flying over a river following a motor boat cruising at a constant speed of 48 km h^{-1} flew at an average altitude of 1 m above the water, initially alone at a distance from the boat ('alone at 1 m'), but then joining other birds in formation ('in formation'). Numbering of circles indicates the position in the formation, '1' being the leader. In formation flights, only birds in or behind the third position were used to measure heart rate, to avoid the possible effects of motor-boat turbulence on flight pattern. Wing-beat frequency of the lead and second birds was measured in groups flying far from the boat. Birds flying with an ultralight aeroplane cruising at $55\text{--}60 \text{ km h}^{-1}$ at an altitude of about 50 m were not able or willing to fly in the wake of the aircraft, and generally flew at a distance, using a flapping flight, at a speed of $45\text{--}50 \text{ km h}^{-1}$ ('alone at 50 m'). Pelicans following the plane finished their flight of 3 km with a glide of 2–3 min ('gliding') before landing on the water. Heart rates averaged 77.6 ± 15.2 beats per min (b.p.m.) when resting, and reached 198.2 ± 18.7 and 204.9 ± 15.7 b.p.m. when paddling in water and when walking, respectively.

the aerodynamic advantage of flight formation⁷, thus reducing the energy saving⁸ — so geese, for example, may make an energy saving of only 2.4% (ref. 8).

In our study, pelicans often had difficulty staying within the formation, particularly when flying at the rear. But even though these birds were regularly adjusting their position, they still achieved a significant energy saving. This saving may be only partly due to effects of the wakes of other birds on the power input that results from formation flight itself. When flying in formation, pelicans appear to beat their wings less frequently and to glide for longer periods. A rough calculation based on our estimates of the proportion of time spent flapping and gliding in formation, and assuming that the overall costs of the glide-flap sequence is the sum of the gliding and flapping components, reveals an actual saving of 1.7–3.4% as a result of wake effects on power input — this value is comparable to that estimated for geese⁸. The main benefit of flight formation, which until now has not been recognized, could be that by flying in a vortex wake, pelicans are able to glide for a greater proportion of their total flight time, with the total energy savings of 11.4–14.0% being achieved primarily through this strategy.

Formation flight, by allowing birds to reduce their energy expenditure, enables them to increase their foraging or migratory range. Oblique and V formations are comparable in terms of energy gain for individual birds apart from the leader, and are aerodynamically the most advantageous compared with other types of flight in flocks⁹. Pelicans use such flight patterns extensively, not only during migration but also during group commuting trips between colonies and foraging zones¹⁰.

Although pelicans benefit by saving

energy when flying in formation, this arrangement may also favour communication and coordination within the group — for example, by helping birds to stay in visual contact^{4,10} and enabling flight-pass and velocity information to be conveyed between them⁶. This may explain why several other naturally occurring configurations of bird flocks are aerodynamically neutral or even disadvantageous relative to solitary flight^{8,11}.

Henri Weimerskirch*, Julien Martin*, Yannick Clerquin†, Peggy Alexandre†, Sarka Jiraskova†

*Centre d'Etudes Biologiques de Chizé,
Centre National de la Recherche Scientifique,
79360 Villiers en Bois, France

e-mail: henriw@cebc.cnrs.fr

†Galatée Films, Le Peuple Migrateur — Jacques Perrin, 92 Rue Joffroy d'Abbas,
75017 Paris, France

1. Lissaman, P. B. S. & Shollenberg, C. A. *Science* **168**, 1003–1005 (1970).
2. Badgerow, J. P. & Hainsworth, F. R. *J. Theor. Biol.* **93**, 41–52 (1981).
3. Hummel, D. J. *Theor. Biol.* **104**, 321–347 (1983).
4. Hamilton, W. J. in *Animal Orientation and Navigation* (ed. Storm, R. M.) 57–71 (Oregon State Univ. Press, Corvallis, 1967).
5. Weimerskirch, H. et al. *Proc. R. Soc. Lond. B* **267**, 1869–1874 (2000).
6. Hainsworth, F. R. *J. Exp. Biol.* **135**, 431–444 (1988).
7. Badgerow, J. P. *Auk* **105**, 749–755 (1988).
8. Cutts, C. J. & Speakman, J. R. *J. Exp. Biol.* **189**, 251–261 (1999).
9. Higdon, J. J. L. & Corrin, S. *Am. Nat.* **112**, 727–744 (1978).
10. O'Malley, J. B. E. & Evans, R. M. *Can. J. Zool.* **60**, 1388–1396 (1982).
11. Gould, L. L. & Heppner, F. *Auk* **91**, 494–506 (1974).

Deep-sea ecology

Developmental arrest in vent worm embryos

Temperature is a key factor in controlling the distribution of marine organisms and is particularly important at hydrothermal vents, where steep thermal gradients are present over a scale of centimetres¹. The thermophilic worm *Alvinella pompejana*, which is found at the vents of the East Pacific Rise (2,500-m depth), has an unusually broad thermotolerance (20–80 °C) as an adult^{2,3}, but we show here that the temperature range required by the developing embryo is very different from that tolerated by adults. Our results indicate that early embryos may disperse through cold abyssal water in a state of developmental arrest, completing their development only when they encounter water that is warm enough for their growth and survival.

We obtained early embryos of *A. pompejana* by *in vitro* fertilization, and reared them at temperatures ranging from 2 °C to 20 °C under atmospheric and deep-sea pressures. We monitored mortality (diagnosed by the breakdown of the plasma membrane or by production of irregular cytoplasmic blebs) and zygotic cleavage during early development.

Embryos kept at 20 °C and one atmosphere of pressure all died within 24 h (Fig. 1a), although many completed the first cleavage. At 14 °C and 10 °C, 70–90% of zygotes cleaved, with rates varying as a function of temperature. At 2 °C, oocytes and embryos remained intact, without cleaving, for 72 h (Fig. 1a) and for at least a further 8 days, when we stopped the experiment.

As low hydrostatic pressure inhibits cleavage in other deep-sea invertebrates⁴, some irregular cleavages in our one-atmosphere incubations were not unexpected. When we repeated the incubation experiments at *in situ* pressures (250 atmospheres),

we obtained qualitatively similar results (Fig. 1b), with fewer abnormalities.

To investigate whether cold-water developmental arrest is reversible, we maintained zygotes at 2 °C for 72 h, exposed them to a 10 °C heat pulse for 45 min, and incubated them at 2 °C for a further 24 h. Although embryos maintained at 2 °C never underwent cleavage, those exposed to a short heat pulse resumed development, and cleavage continued even after the embryos were

moved back into cold water (Fig. 1c).

The large difference in temperature tolerance between adults (20–80 °C) and embryos (2–20 °C) precludes the possibility of embryonic development inside adult worm tubes. Optimal temperatures for development (10–15 °C) are found close to the bases of hydrothermal-vent chimneys. As eggs are negatively buoyant upon release, some embryos probably develop on or near the sea floor, just below the adult habitat. However, it is likely that at least some embryos are dispersed by currents and carried to new sites through cold (2 °C) abyssal sea water, as occurs in other vent species^{5,6}. Our results show that embryos of *A. pompejana* survive but do not develop at this temperature.

It has been suggested that dispersing larvae of hydrothermal-vent bivalves⁷ and polychaetes⁸ may delay their development until they encounter warm water. Our results provide empirical evidence for such reversible developmental arrest in a vent species; a similar phenomenon has been reported in larvae of the bathyal echinoid *Linopneustes longispinus*⁹. Although we do not know how long the embryos of *A. pompejana* remain viable at low temperatures, this temperature-sensitive mechanism for controlling development may result in

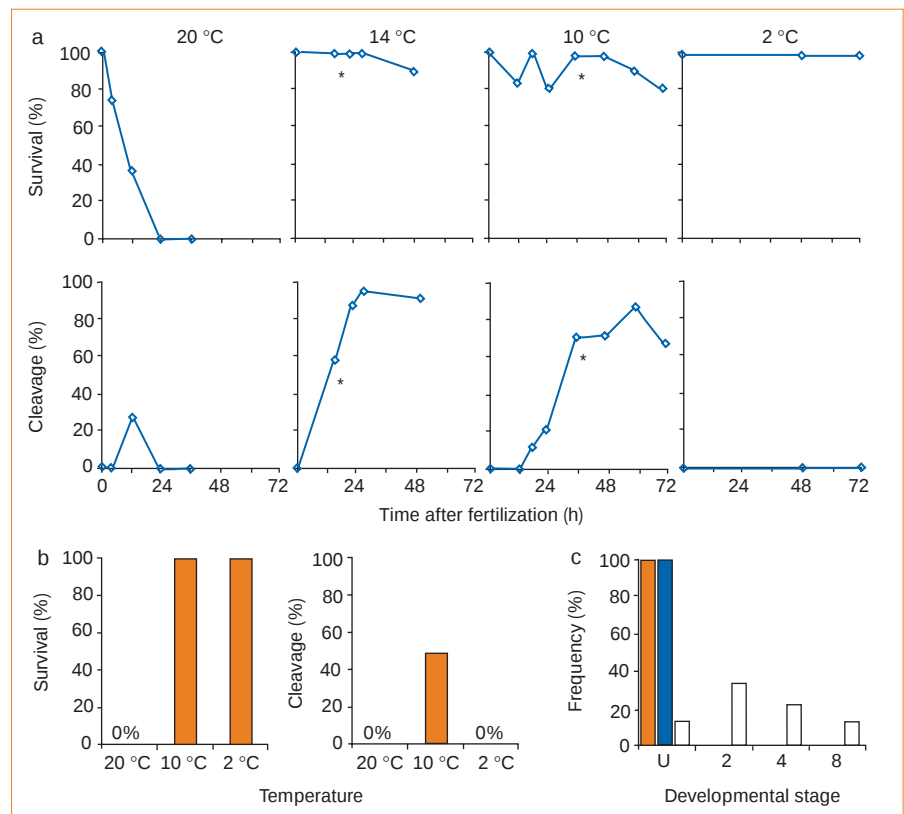


Figure 1 Effects of temperature on early embryos of *Alvinella pompejana*. **a**, Percentage of embryos surviving (top) and cleaving (bottom) when incubated at one atmosphere of pressure and at temperatures of 2–20 °C. At least 20 embryos were scored per sample, except for 2 samples (asterisks), which contained fewer embryos. **b**, Survival and cleavage at the indicated temperatures after 48 h at a pressure of 250 atmospheres. **c**, Distribution of cleavage stages in embryos incubated at 2 °C for 24 h (orange bar) and 8 days (blue bar) and in 72-h embryos exposed to a 45-min heat pulse at 10 °C (white bars), then transferred back to 2 °C for a further 24 h. Developmental stages are two- to eight-cell stages. U, uncleaved.

very long dispersal distances.

**Florence Pradillon*, Bruce Shillito*,
Craig M. Young†, Françoise Gaill***

*Biologie Marine, Unite Mixte de Recherche,
Université Pierre et Marie Curie, 7 Quai
Saint-Bernard, 75252 Paris cedex 05, France
e-mail: francoise.gaill@snv.jussieu.fr

†Division of Marine Science, Harbor Branch
Oceanographic Institution, 5600 US Highway 1
North, Fort Pierce, Florida 34946, USA

1. Edmond, J. M., Von Damm, K. L., McDuff, R. E. & Measures, C. I. *Nature* **297**, 187–191 (1982).
2. Chevaldonné, P., Desbruyères, D. & Childress, J. J. *Nature* **359**, 593–594 (1992).
3. Cary, C. S., Shank, T. & Stein, J. L. *Nature* **391**, 545–546 (1998).
4. Young, C. M. & Tyler, P. A. *Limnol. Oceanogr.* **38**, 178–181 (1993).
5. Mullineaux, L. S. & France, S. C. in *Seafloor Hydrothermal Systems: Physical, Chemical, Biological and Geological Interactions* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 408–424 (Am. Geophys. Un., Washington DC, 1995).
6. Marsh, A. G., Mullineaux, L. S., Young, C. M. & Manahan, D. T. *Nature* **411**, 77–80 (2001).
7. Lutz, R. A., Jablonski, D., Rhoads, D. C. & Turner, R. D. *Mar. Biol.* **57**, 127–133 (1980).
8. Zottoli, R. A. *Proc. Biol. Soc. Wash.* **96**, 379–391 (1983).
9. Young, C. M. & Cameron, J. L. in *Reproduction, Genetics and Distributions of Marine Organisms* (eds Ryland, J. S. & Tyler, P. A.) 225–231 (Olsen & Olsen, Fredensborg, Denmark, 1989).

Geophysics

Longitudinal variation in springtime ozone trends

Satellite measurements indicate that mid-latitude ozone depletion in the Northern Hemisphere spring during 1979–97 was worst over Europe and Russia¹. Here we show that these longitudinal differences in ozone trends are due to a combination of decadal variations in the circulation^{2,3} and transport of ozone-depleted air from the polar vortex. Any increase in ozone depletion in the polar vortex as a result of future cooling of the stratosphere would therefore be particularly bad over Europe and Russia.

Massive ozone depletion inside the Arctic vortex has occurred frequently during the past decade. As the total ozone mapping spectrometer satellite data we use here have gaps from December 1994 to July 1996 and in May 1994 (ref. 4), the principal part of the vortex depletion during 1979–97 is included by using the 1993 and 1997 depletions (depletions originating from the vortex in 1997 are taken from ref. 5). They were calculated by tracking the ozone-depleted air with reverse domain-filling trajectory calculations covering the altitude region of ozone depletion.

We calculated diabatic descent using a radiation scheme⁶; ozone-depletion mixing ratios were regridded every seventh day to introduce almost realistic average mixing⁵. The 1993 calculations were carried out in the same way, and the vortex depletion was

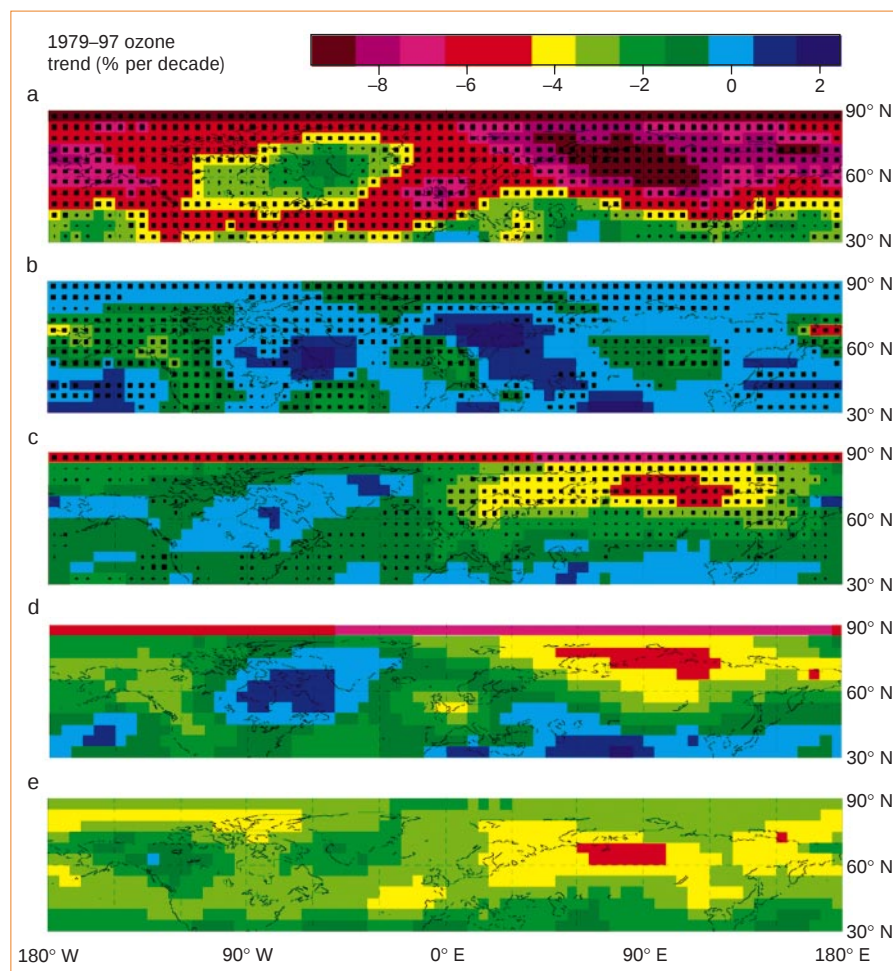


Figure 1 Ozone trends in the April–May periods of 1979–97. **a**, Calculated overall ozone trend. Large black dots indicate a significant trend with 95% confidence; small dots indicate 68% confidence. **b**, Ozone trend due to the trend in the height of 25,000-pascal pressure. This explains why the downward ozone trend is small south of Greenland and large over the United Kingdom. **c**, Ozone trend due to vortex depletions in 1993 and 1997. This accounts for most of the large downward ozone trend over Scandinavia and Russia. **d**, Sum of the trends shown in **b** and **c**. This explains 81% of the variance in mid-latitude ozone trends in **a**. **e**, The residual trend, caused by factors such as local ozone depletion outside the vortex, is roughly independent of longitude.

taken from ref. 7, using a correction for diffusive transport into the vortex taken from 1997 calculations⁸. The column-integrated depletion amounts to 79 Dobson units, which is comparable to earlier values⁹.

We applied a multiple linear-regression model¹ to the column ozone, assuming a linear trend and including the quasi-biennial oscillation (with appropriate lag¹⁰) and solar cycle as independent variables. The trend is shown in Fig. 1a: black dots indicate where trends are significant. To calculate the effect of circulation changes and vortex depletions on the ozone trends, we used the geopotential height at a pressure of 25,000 pascals (geometric height, about 10 km) supplied by the European Centre for Medium-Range Weather Forecasts, as well as the transported depletion in 1993 and 1997 (zero in other years), instead of the linear ozone trend in the regression model.

Figure 1b shows the trend in ozone levels that may be explained by the trend in the 25,000-pascal geopotential height. The height trend accounts for the fact that the

downward ozone trend is small south of Greenland and large over the United Kingdom, but it does not explain the large trends over Scandinavia and Russia, for example.

Figure 1c shows the effect of vortex depletions on ozone trends and helps to explain the large downward ozone trends over Scandinavia and Russia. The positive trends shown are non-physical and not significant. By neglecting vortex depletion in years other than 1993 and 1997, the effect of vortex depletion will be underestimated. Statistics showing the April–May location of the vortex (remnants) in two other years with large vortex ozone depletions (1996 and 2000) reveal that the most probable location is over Europe and Russia, indicating that this is a robust, climatological feature (data not shown). The part of the mid-latitude (30°–60° N) trend that may be attributed to vortex depletions is 0.8% per decade, or 19% of the observed ozone trend on average. This might be compared to the 25% of the observed ozone depletion in

April–May 1997 that was found to originate from vortex depletions⁵.

Figure 1d incorporates the trends in Fig. 1b, c, and this combination accounts for a large fraction of the longitudinal trend variations. From 30°–60° N, it explains 81% of the variance (correlation, 0.90), whereas the height trend alone explains only 35% of depletion during the April–May periods of 1979–97. Figure 1e shows the trends of the residuals from the latter regression. The remaining downward trend is about 3% per decade and is caused by factors such as local ozone depletion outside the vortex. The remaining longitudinal variation could be due to failure to take vortex depletions in other years into account.

In the Southern Hemisphere, the vortex is much more circumpolar and breaks up about two months later in the season than in the Northern Hemisphere. Depletion of the Antarctic vortex therefore has little influence on southern mid-latitudes in spring, and this helps to explain why the downward ozone trend in northern mid-latitudes is much larger than in their southern counterparts in spring¹.

Bjørn M. Knudsen, Signe B. Andersen

Danish Meteorological Institute, Lyngbyvej 100,
2100 Copenhagen, Denmark
e-mail: bk@dmu.dk

1. World Meteorological Organization *Glob. Ozone Res. Monitor. Proj. Rep.* **44**, Ch. 4 (1999).
2. Hood, L. L., Rossi, S. & Beulen, M. *J. Geophys. Res.* **104**, 23321–24339 (1999).
3. Entzian, G. & Peters, D. *Proc. Quad. Ozone Symp.*, Sapporo, 3–8 July 2000.
4. McPeters, R. D., Hollandsworth, S. M., Flynn, L. E., Herman, J. R. & Seftor, C. *J. Geophys. Res. Lett.* **23**, 3699–3702 (1996).
5. Knudsen, B. M. & Groop, J.-U. *J. Geophys. Res.* **105**, 6885–6890 (2000).
6. Morcrette, J.-J. *J. Geophys. Res.* **96**, 9121–9132 (1991).
7. Larsen, N., Knudsen, B., Mikkelsen, I. S., Jørgensen, T. S. & Eriksen, P. *Geophys. Res. Lett.* **21**, 1611–1614 (1994).
8. Knudsen, B. M. *et al.* *Geophys. Res. Lett.* **25**, 627–630 (1998).
9. Müller, R. *et al.* *Nature* **389**, 709–712 (1997).
10. Zerefos, C. S., Bais, A. F., Ziomias, I. C. & Bjokov, R. D. *J. Geophys. Res.* **97**, 10135–10144 (1992).

Plant genetics

Ancient wild olives in Mediterranean forests

Early domestication and extensive cultivation¹ have meant that staple Mediterranean fruit crops such as olives, grapes and dates exist in wild-looking forms that are secondary derivatives produced by sexual reproduction among cultivated plants (cultivars), which were initially propagated vegetatively². By using genetic markers associated with characters that render plants unsuitable for domestication, we show here that genuinely wild olive trees, which cannot be distinguished morphologically from feral forms, still survive in a few Mediterranean forests. These wild stocks are genetically distinct

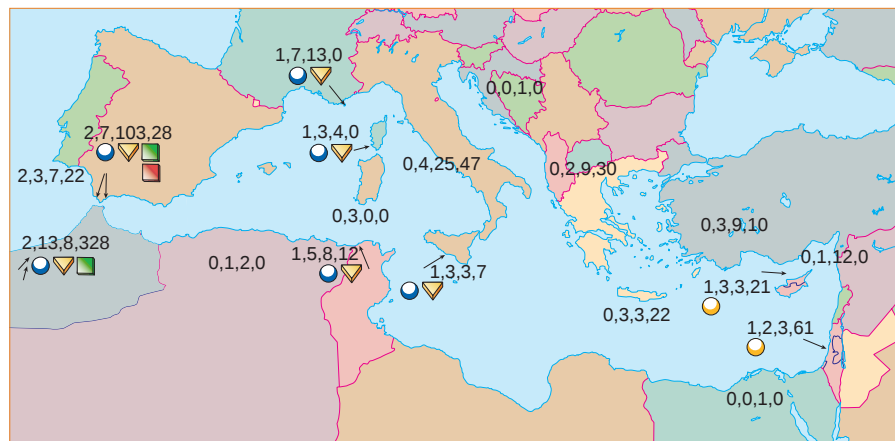


Figure 1 Geographical distribution of forests tested for the presence of genuinely wild olive trees. Five different alleles (each represented by a different symbol) peculiar to these trees were found to be present at frequencies of 4–32%, 8–14%, 20–24%, 10–22% and 4–15%, respectively; these were present exclusively in the 10 forests suspected of containing genuinely wild olives. Arrows, population locations tested. For each country, the sequence of four numbers indicates populations of wild oleasters, feral oleasters, formal cultivars and locally cultivated clones, respectively, that were scored genetically.

and more variable than either the crop strains or their derived feral forms, a finding that has important implications for the conservation of these ancient lineages.

We surveyed ten forests in seven countries around the Mediterranean basin (Fig. 1) to identify any surviving genuinely wild olive trees. These forests were selected as the most likely to contain these rare trees, according to whether they were exposed to past and present climatic conditions suitable for oleaster growth³, and from consideration of their past and present isolation from areas of cultivated olive trees. We also assessed whether the large oleaster populations (wild and feral forms are collectively known as oleasters) in these forests might have offset the influence of occasional pollen and seed flow from cultivated areas. In the eastern part of the basin, where olive trees have been extensively cultivated for longer periods¹, we used less restrictive criteria, but still found no candidate forests in Egypt, Syria, Turkey, Crete or Greece.

In each selected forest, we sampled 40 trees and scored them for allozyme markers, particularly at loci that are known to be genetically linked to regions that control the juvenile phase⁴ which, if prolonged, is unsuitable for olive domestication. We compared our results to those obtained at the same loci from 802 olive clones representing the main domesticated olive^{5–7} and from 1,395 feral olives from 62 sites^{4,8}, including forests throughout the Mediterranean basin that did not satisfy our selection criteria.

Of the 26 alleles identified at eight loci, two alleles at each of the two loci associated with characters unsuitable for domestication, as well as another allele at another locus, were present exclusively in the populations of the 10 selected forests considered most likely to contain genuinely wild olives

(Fig. 1). The other alleles were common to cultivated olives and oleasters, whatever their origins. In the occidental area, the average genetic differentiation values (F_{st}) are 6% between the eight putative wild populations and 22% between these and other nearby oleaster populations.

Our results provide evidence of the survival of indigenous oleaster populations, particularly in the western part of the basin, as suggested previously¹. Genetic-diversity values over cultivars, feral olives and the wild olives of the selected forests were 0.286, 0.414 and 0.506, which is consistent with our interpretation that the domesticated olive represents a sample of the genetic variation in genuinely wild olive populations that persist today. Owing to their very long lifespan, these wild trees should be closely related to the neolithic olives recognized as the crop progenitor. Our discovery will encourage the preservation of environmental conditions that favour the survival of large populations of oleasters that are properly isolated from cultivated olive trees.

Roselyne Lumaret*, Noureddine Ouazzani†

*Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de la Recherche Scientifique, 1919 Route de Mende, F-34293 Montpellier cedex 5, France
e-mail: lumaret@cefe.cnrs-mop.fr

†Ecole Nationale d'Agriculture, BP S40, Meknès, Morocco

1. Zohary, D. & Spiegel-Roy, P. *Science* **187**, 319–327 (1975).
2. Turill, W. B. *Kew Bull.* **3**, 439–442 (1951).
3. Terral, J. F. & Mengüel, X. *Palaeogeogr. Palaeoclim. Palaeoecol.* **153**, 71–92 (1999).
4. Ouazzani, N., Lumaret, R., Villemur, P. & Di Giusto, F. *J. Hered.* **84**, 34–42 (1993).
5. Ouazzani, N., Lumaret, R. & Villemur, P. *Agronomie* **15**, 31–37 (1995).
6. Trujillo, I., Rallo, L. & Arùs, P. *J. Am. Soc. Hort. Sci.* **120**, 318–324 (1995).
7. Ouazzani, N., Lumaret, R. & Villemur, P. *Euphytica* **91**, 9–20 (1996).
8. Lumaret, R., Ouazzani, N., Michaud, H. & Villemur, P. *Bocconea* **7**, 39–42 (1997).

Splicing-related catalysis by protein-free snRNAs

Saba Valadkhan & James L. Manley

Department of Biological Sciences, Sherman Fairchild Center of Life Sciences, Columbia University, New York, New York 10027, USA

Removal of intervening sequences from eukaryotic messenger RNA precursors is carried out by the spliceosome, a complex assembly of five small nuclear RNAs (snRNAs) and a large number of proteins. Although it has been suggested that the spliceosome might be an RNA enzyme, direct evidence for this has been lacking, and the identity of the catalytic domain of the spliceosome is unknown. Here we show that a protein-free complex of two snRNAs, U2 and U6, can bind and position a small RNA containing the sequence of the intron branch site, and activate the branch adenosine to attack a catalytically critical domain of U6 in a reaction that is related to the first step of splicing. Our data provide direct evidence for the catalytic potential of spliceosomal snRNAs.

An essential step in the maturation of eukaryotic pre-mRNA is the removal of intervening sequences (introns) from the coding sequences (exons). The spliceosome—a dynamic assembly of five snRNAs and a large number of proteins—catalyses intron removal, or splicing, through two consecutive transesterification reactions^{1,2}. In the first step, the 2'OH group of a conserved adenosine residue, bulged from an RNA–RNA duplex, is activated to attack the phosphodiester bond at the 5' splice site, resulting in the formation of a branched (lariat) intron containing a 2' to 5' phosphodiester bond and release of the 5' exon. The second step involves a nucleophilic attack by the 3'OH of the 5' exon on the phosphodiester bond at the 3' splice site, resulting in ligation of the two exons and excision of the intron lariat. Complementation studies *in vivo* and *in vitro* in addition to fundamental similarities to self-splicing group II introns^{3–8} have suggested that the spliceosome might be an RNA enzyme. Of the five spliceosomal snRNAs, U2 (which base-pairs with the intron branch site) and U6 (which interacts with the 5' splice site) are the only ones required for both steps of splicing^{9,10}, and mutagenesis studies support possible roles in catalysis for these two RNAs^{11–17}. However, the ability of these RNAs and/or spliceosomal proteins^{18,19} to participate directly in catalysis has not been demonstrated. Previous studies that investigated the possible catalytic activity of snRNAs used *in vitro* evolution and selection^{20,21} and isolated U6 and/or U2-derived ribozymes that catalysed reactions not closely related to the mRNA splicing reactions.

We have previously shown that U2 and U6 snRNAs can form a base-paired complex in the absence of proteins that closely resembles the complex thought to exist in the active spliceosome, and we also provided evidence for the presence of a genetically proven tertiary interaction²² in this complex (Fig. 1, ref. 23, and S.V. and J.L.M., unpublished data). We show here that this complex can bind and position a short RNA oligonucleotide that contains the branch consensus sequence, and catalyses a reaction in which the 2'OH group of a bulged adenosine attacks a specific phosphodiester bond in U6, resulting in the formation of an unusual phosphotriester bond. The sequence and ionic requirements for this reaction closely resemble those of authentic splicing. Our data provide evidence for the ability of spliceosomal snRNAs to function as the catalytic domain of the spliceosome.

Catalytic activity of the U2–U6 complex

To assay the possible catalytic activity of spliceosomal snRNAs in the absence of proteins, we incubated a previously characterized U2–U6 snRNA complex, consisting of *in vitro* transcribed and purified

segments of human U2 and U6 snRNAs²³, with radiolabelled short RNA oligonucleotides that contained the consensus branch site sequence (the branch oligonucleotide (Br), Fig. 1) or the 5' splice-site consensus sequence under various ionic, temperature and pH conditions. Although no significant reaction was observed with the 5' splice site oligonucleotide, in the presence of Mg²⁺ and after 24 h incubation at room temperature about 0.1% of Br was converted to a new, low-mobility RNA species on denaturing polyacrylamide gel electrophoresis (PAGE) (Fig. 2a). Formation of this product (RNA X) was dependent on the presence of both U2 and U6 in the reaction, and was greatly reduced after incubation on ice or at temperatures close to the melting temperature (*T*_m) of the U2–U6 complex (not shown). RNA X accumulated linearly for at least 2 h and continued to increase for nearly 20 h (Fig. 2b). Varying the concentration of Br or the U2–U6 complex over a wide range resulted in roughly linear increases in the formation of RNA X, followed by a plateau at higher concentrations (Fig. 2c and data not shown), with an apparent Michaelis constant (*K*_m) of 5 μM and a reaction rate (*k*_{obs}) of 0.002 min^{–1} under standard reaction conditions. To prove that RNA X indeed resulted from a new covalent linkage and not an unusually strong non-covalent interaction, purified RNA X was denatured at 90 °C for 5 min in 9 M urea plus

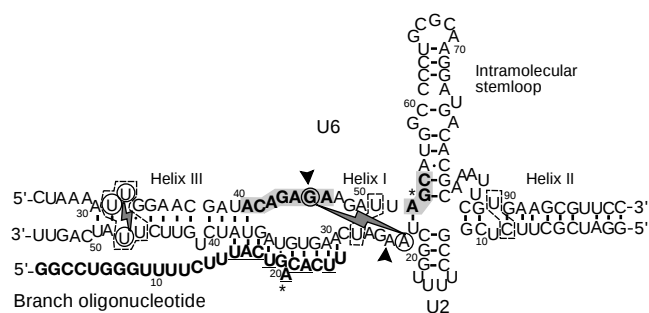


Figure 1 Base-pairing interactions in the *in vitro*-assembled complex of U2–U6 and the branch oligonucleotide (Br). Shaded boxes mark the invariant regions in U6 and previously established base-paired regions are indicated. Dashed lines connect psoralen-crosslinkable nucleotides (S.V. and J.L.M., unpublished data). The circled residues connected by a zigzag can be crosslinked by ultraviolet light. The underlined residues in Br constitute the yeast branch consensus sequence. Asterisks denote the residues involved in the covalent link between Br and U6 in RNA X (see text). Arrowheads point to residues involved in a genetically proven interaction in yeast²². Numbers indicate nucleotide positions from the 5' ends of full-length human U2 and U6.

12 mM EDTA, rapidly cooled and loaded on a gel containing 8 M urea and 40% formamide. No detectable dissociation was observed (Fig. 2d). Notably, RNA X showed anomalous mobility on 8% and 16% denaturing gels compared with RNA markers of known length, which is a property indicative of nonlinear RNA species (Fig. 2e).

We next tested the ionic requirements for the formation of RNA X (Fig. 2f and data not shown). RNA X formed with the same efficiency in buffers that contained HEPES, MOPS or sodium cacodylate instead of Tris. Increasing the Mg^{2+} concentration from 10 to 200 mM had a modest effect on RNA X formation, whereas RNA X did not form in the absence of divalent cations or in the presence of 0.4 M NaCl, which fully supports the formation of the annealed U2–U6 complex²³. Notably, RNA X did not form with Mn^{2+} as the only divalent cation in the buffer. Indeed, addition of Ca^{2+} , Mn^{2+} or ammonium salts to Mg^{2+} reduced the efficiency of RNA X formation, although Ca^{2+} alone weakly supported product formation.

The bulged adenosine in Br attacks the AGC triad of U6

To determine the constituents of RNA X, reactions were performed with 5' or 3' end-labelled U2, U6 or Br. Label at either end of U6 or Br, but not U2, appeared in RNA X (not shown). These observations, which are consistent with an RNase T1 fingerprint of 5' end-labelled RNA X (not shown) and with results from alkali treatment of the product (see below), indicate the presence of full-length U6 and Br, but not U2, in RNA X. Dephosphorylation of the 5' end or introduction of a cyclic phosphate at the 3' end of U6 or Br did not affect the rate of RNA X formation (not shown), indicating that the 5' and 3' ends of U6 and Br are not involved in the new covalent bond in RNA X. These results suggest that RNA X is probably an X-shaped molecule.

To map the location in U6 of the covalent bond between U6 and Br, purified RNA X was annealed to an oligonucleotide that was complementary to the 3' end of U6 (nucleotides 81–99) and subjected to reverse transcription. Notably, a strong stop immediately 3' to A53, with a minor stop 5' to that residue, was observed (Fig. 3a). This is significant because A53 lies within the so-called AGC triad, which is one of two invariant, functionally critical regions of U6 (refs 1, 2, 11–14), and which is also conserved in U6atac (the U6 counterpart in the ATAC spliceosome)^{1,24,25}. Base or backbone mutations in this region can be frequently incompatible with splicing^{1,11–17}. A similarly positioned AGC triad in domain 5 of

self-splicing group II introns is thought to be part of the active site of that ribozyme^{1,5}.

We next used limited enzymatic digestion with RNase T2 (Fig. 3b), T1 (Fig. 3c) and P1 (not shown) of purified RNA X labelled at the 5' end of Br to map the location of the linkage in Br. Whereas scission of phosphodiester bonds up to G20 resulted in release of 5' fragments of Br, fragments resulting from digestion after A21 were absent (Fig. 3b, c and data not shown), suggesting that this residue is involved in the covalent linkage between U6 and Br (see below). The results obtained from Br point mutants that contained RNase T1 digestion sites close to A21 (Br C18G, Br U19G and Br A23G, which have at most small effects on catalysis; see below) confirmed the data from wild-type Br (Fig. 3c and data not shown). The above data also rule out the possibility that RNA X is a hyperstable, non-covalent complex by demonstrating its accessibility to RNases (and iodoethanol cleavage; see below) throughout its length. Taken together, the mapping data are consistent with the hypothesis that RNA X is X-shaped, with U6 and Br joined covalently through residue A53 in the conserved AGC triad in U6 and the bulged A (A21; see below) in Br.

A phosphotriester linkage between U6 and Br

We next wished to characterize the nature of the linkage between U6 and Br in RNA X. We first subjected internally labelled RNA X to complete RNase P1 or T2 digestion, but were unable to isolate a nuclease-resistant core (data not shown). This indicates that the covalent bond between U6 and Br can either be digested by these enzymes or is susceptible to hydrolysis mediated by solvent exposure after complete digestion (see below). In any event, we required alternative methods to characterize the U6–Br linkage.

Treatment of phosphorothioate-substituted RNA with iodoethanol results in specific backbone cleavage at sites of phosphorothioate substitutions, and can thus be used to map sites of interaction or interference that prevent cleavage²⁶. We therefore used iodoethanol cleavage with 5' and 3' end-labelled RNA X prepared from phosphorothioate-substituted U6, in which each U6 contained, on average, one randomly distributed phosphorothioate linkage²⁷ (see Methods). The iodoethanol cleavage patterns confirmed the reverse transcription mapping data and, more importantly, localized the site of covalent bond formation to the backbone phosphate between A53 and G54 (Fig. 4a). In α -S-GTP-substituted U6 RNA, the iodoethanol cleavage ladders stopped at G54 with both 5'- and

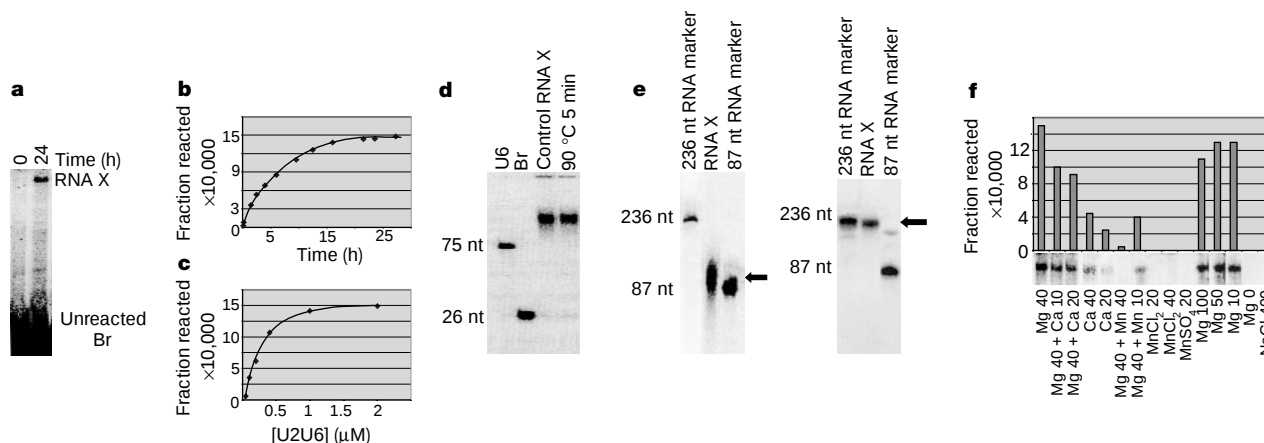


Figure 2 Characterization of RNA X. **a**, Formation of RNA X. Reaction products were resolved by denaturing PAGE, and the gel is shown from the level of the wells to the bottom. **b**, Time course of formation of RNA X. **c**, RNA X formation as a function of U2–U6 complex concentration [U2U6]. **d**, Resistance to denaturation. Control RNA X, RNA X after two gel-purification steps, before stringent denaturation. The denaturation condition is indicated above the far right lane. Gel composition: 16% PAGE, 8 M urea and 40%

formamide. Sizes of U6 and Br are indicated. nt, nucleotide. **e**, Anomalous gel mobility of RNA X. The arrows point to the location of RNA X. Gel composition: 8% (left) and 16% (right) denaturing PAGE. **f**, Cationic requirements for formation of RNA X. The concentration of cations in each reaction (in mM) is indicated below each lane. Efficiency of RNA X formation was quantified by PhosphorImager and plotted.

3'-labelled U6 (Fig. 4a, left and right panels, respectively). This contrasted with results obtained from iodoethanol cleavage of phosphorothioate substitution at other sites (Fig. 4a) and with other α -S-nucleoside 5'-triphosphate (NTP) substitutions (for example, Fig. 4b) in which one of the two U6 fragments remained attached to Br, and thus was not detected at its normal location. For example, cleavage of the phosphorothioate bond 5' of G49 resulted in a U6 25–48 fragment that is released from RNA X (Fig. 4a, left panel) and a U6 49–95 fragment that stays attached to Br and

therefore has a much slower mobility—such that it was not resolved on the gels (Fig. 4a, right panel)—compared with the U6 49–95 fragment in the control lane. The unusual iodoethanol cleavage pattern with G54 phosphorothioate substitution can be explained if the link between U6 and Br is such that iodoethanol cleavage at the phosphorothioate 5' of G54 would break U6 in two halves such that either half of U6 can be released from RNA X, that is, if the link between U6 and Br in RNA X is on the phosphorus atom (Fig. 4c). Indeed, the reaction of iodoethanol with a phosphorothioate triester results in release of either of the three phosphate ligands²⁸. A covalent bond originating on a sugar or a base in U6 would have resulted in an iodoethanol cleavage pattern in which one of the two U6 halves would remain attached to Br. Involvement of backbone functionalities as reacting groups in RNA X is consistent with the mild phenotype of U6 mutations in the AGC region (see below).

To extend this analysis, we took advantage of the alkali sensitivity of RNA X (half-life in pH 12 at 50 °C: <3.5 min). Purified RNA X labelled at the 5' end of U6 or Br was treated with alkali (pH 12, 20 min at room temperature) and compared with similarly labelled U6 or Br. These conditions resulted in partial conversion of RNA X to full-length U6 and Br without significant backbone phosphodiester breaks, confirming that the chemical bond formed in the product is unusually alkali-sensitive (Fig. 4d, lane 1). No difference in mobility of released U6 and Br compared with unreacted controls could be observed, even after long electrophoresis times (not shown). In addition, alkali hydrolysis of U6 and Br released from RNA X resulted in fragments with identical mobilities to those of controls (Fig. 4d, compare lanes 4 and 7 with 5 and 6), even though a difference in relative molecular mass of less than 40 could have been detected (for example, compare the two hydrolysis ladders in Fig. 3c). As formation of an adduct between U6 and Br without a leaving group is entropically unfavourable, the above results are consistent with a very small leaving group, such as H₂O. Taken together, the alkali sensitivity of RNA X and the phosphorothioate substitution data are consistent with a phosphotriester linkage between U6 and Br in RNA X, with the phosphate bond between A53 and G54 forming the triester.

Sequence requirements for RNA X formation

We next examined the effects of mutations in Br and U6 on RNA X formation (Fig. 5). Mutations that disrupted the base-pairing interaction between U2 and Br (Fig. 5a, transversion mutation) or that changed the base identity of A21 (A21 to G or C) resulted in marked decreases in the formation of RNA X. Mutations that removed the bulged residue in Br (deletion of A21 or G20) or disrupted U2–U6 helix III by competing with U6 for base-pairing to U2 in this region (Fig. 5a, 5'-complementary mutant) had the most marked phenotype, reducing the rate of RNA X formation essentially to zero. Point mutations in other positions or removal of residues 1–14 of Br had modest effects on the formation of RNA X (Fig. 5a and not shown). Notably, mutations in the ACAGAGA box of U6 (Fig. 5b; see also Fig. 1)—one of the two invariant, functionally critical regions of U6 (refs 1, 2, 11–14, 16, 17)—resulted in loss of catalytic activity, although mutations in the AGC triad had only mild effects. The discrepancy between the modest effect of AGC mutants in our system and stronger phenotypes seen in some cases *in vivo*^{11–15} might reflect a role for the AGC triad other than 2'OH positioning and activation, for example, a function in 5' splice-site coordination. Alternatively, different steps might be rate limiting in the two systems.

All Br mutants were also tested for binding to the U2–U6 complex (see Methods) to distinguish between binding versus catalytic defects. As expected, deletion of Br A21 or G20 and the 5'-complementary mutation resulted in significant increases in binding, whereas other Br mutations, with the exception of the transversion mutation, did not significantly affect binding (not shown). To determine whether the two mutations in the U6

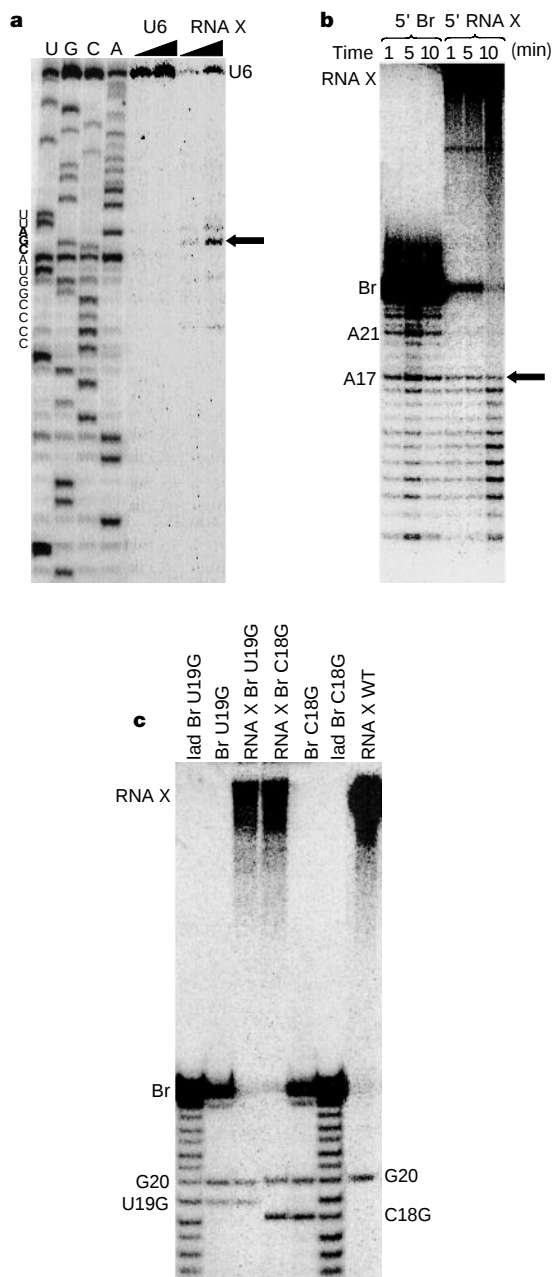


Figure 3 Mapping the site of the covalent linkage in RNA X. **a**, Reverse transcription mapping of the site of the covalent link in U6. Lanes marked U, G, C and A are sequencing reactions. Reverse transcription of two concentrations of U6 and RNA X are presented. The arrow points to the site of the strongest stop. A partial sequence of U6 is indicated to the left of the panel, with the AGC triad in bold letters. **b**, Limited RNase T2 digestion of RNA X labelled at the 5' end of Br. Times of T2 digestion are indicated. The arrow points to the site of stop in the ladder. **c**, Mapping the site of the covalent link on Br by limited RNase T1 digestion of RNA X that was 5'-labelled on Br. The identity of the Br derivative used is indicated above each lane. Lad, alkaline hydrolysis ladder; WT, wild type.

ACAGAGA box interfered with binding or catalysis, we assayed Br binding directly and also tested the activity of the U6 mutants as a function of Br concentration. The results (not shown) indicate that both mutants were defective in catalysis, although the U6 44–47 UCUC mutation, which could enhance base pairing with U2 and

hence competition with Br (see Fig. 1), also decreased binding. Thus, in marked similarity to the splicing reaction, formation of RNA X requires an intact U2–U6–Br base-pairing interaction, the ACAGAG box in U6 and a bulged A.

We were able to partially suppress the loss of function resulting

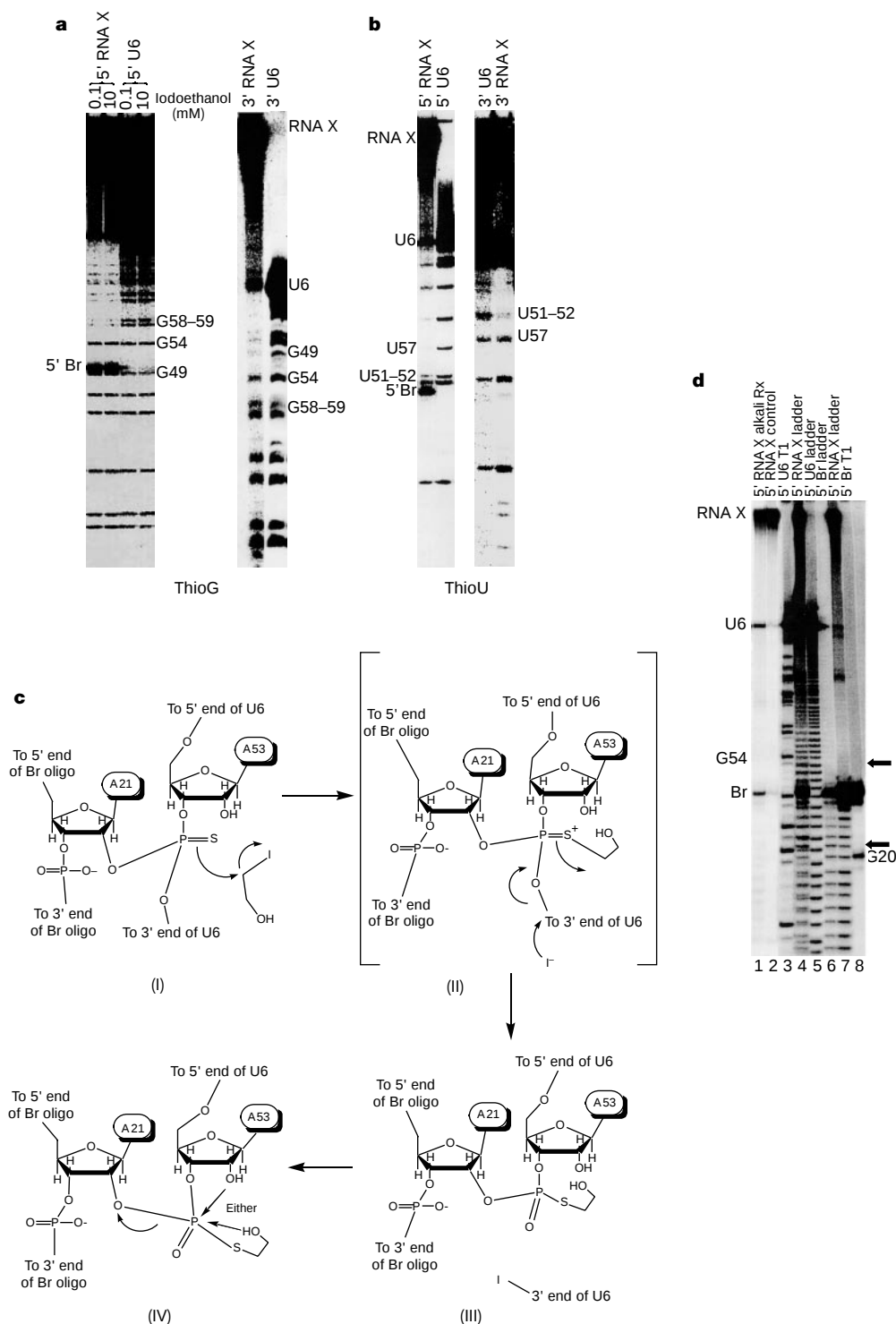


Figure 4 Characterization of the new covalent bond. **a**, Iodoethanol treatment of 5'-labelled (left) and 3'-labelled (right) RNA X formed with α -S-GTP. **b**, Iodoethanol treatment of 5'-labelled (left) and 3'-labelled (right) RNA X formed with α -S-UTP. **c**, The iodoethanol cleavage pathway for the proposed phosphorothioate triester (adapted from ref. 28). Sulphur launches a nucleophilic attack on iodoethanol (I), resulting in alkylation of sulphur (II). In a concerted reaction, iodide ion reacts with either of the phosphorus ligands

(II). The resulting alkyl phosphorothiodiester (III) can further dissociate²⁶ after a nucleophilic attack on the phosphodiester by either the 2'-OH of A53 or the OH of the alkylated sulphur (IV). Oligo, oligonucleotide. **d**, Limited alkali hydrolysis of RNA X. T1, limited RNase T1 digestion reactions; ladder, limited alkali hydrolysis reactions; alkali Rx, mild alkali treatment reaction. Arrows point to the two nucleotides involved in the covalent bond between U6 and Br. 5'-labelled RNA X is labelled on both U6 and Br.

from the A21G mutation in Br by a compensatory mutation in U6 that changed G54 to A (Fig. 5c). This observation is consistent with a direct interaction between A21 and G54, probably as a symmetrical G–A mismatch²⁹ (Fig. 5c). This result further supports the mapping data in U6 and Br, and also shows that the strong effects of A21 mutations result from loss of a functionally important interaction and not the absence of a reacting group. The interaction might serve to recognize and/or position the bulged residue of Br in the catalytic centre. Notably, the functional groups involved in this type of G–A base pair, N1 and N6 of adenosine, have been implicated in branch recognition in the spliceosome³⁰. A similar interaction might be one of the means by which the branch adenosine is recognized and positioned in the active site of the spliceosome.

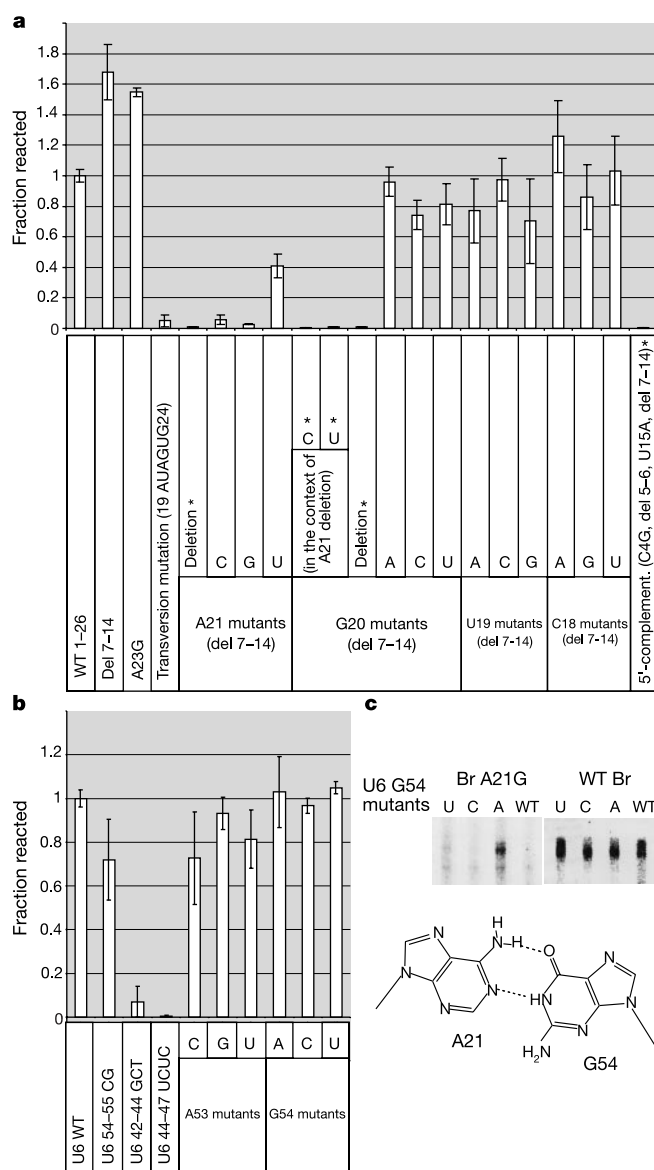


Figure 5 Mutagenesis of the branch oligonucleotide and U6. **a**, Summary of mutations in Br. Complement., complementary. **b**, Summary of mutations in U6. The fraction of reacted Br (RNA X) was arbitrarily set to 1 for wild type (WT). For mutants that show greatly increased binding (lanes marked with asterisks) the data shown represent the apparent k_{cat} of the reaction normalized to the value for wild type, which was arbitrarily set to 1 (see Methods). Lanes without error bars have a standard deviation too small to be visible. **c**, Suppression of the Br A21G mutation by U6 G54A. The model for the symmetrical G–A mismatch base pairing is shown at the bottom.

We next investigated whether the 2'OH on the bulged A in Br (A21) participates in catalysis. A DNA oligonucleotide with identical sequence to Br was not active in RNA X formation (data not shown). Use of chemically synthesized Br with a single 2'H substitution at residue A21 (Br A21 2'H) resulted in a marked reduction in formation of RNA X, although low levels were still detected (Fig. 6a). In the spliceosome, the residue immediately 5' to the branch nucleotide (G20 here) can also assume a bulged conformation and serve as the nucleophile for the first step of splicing³¹, and this could be responsible for the residual RNA X formation. Alternatively, the small amount of product formed with Br A21 2'H could reflect a linkage between Br and U6 located elsewhere in the two molecules³¹. To distinguish between these two possibilities, we tested a Br derivative with 2'H at both A21 and G20 (Br G20–A21 2'H). Formation of RNA X with this double-substituted Br was nearly undetectable, even at very high Br concentrations (Fig. 6a). Although it cannot be ruled out that the effect of 2'H substitution at G20 reflects a general stimulatory effect of a 2'OH group at this position, these results support the idea that at least most of the low Br A21 2'H activity results from G20 functioning as the nucleophile. To rule out the possibility that the inactivity of the deoxyribonucleoside-substituted Br oligonucleotides resulted from a binding defect rather than a defect in chemistry, we tested them both for binding to the U2–U6 complex and for their ability to competitively inhibit the reaction of wild-type Br with the U2–U6 complex. Both Br derivatives behaved indistinguishably from wild-type Br in these assays (data not shown). Together with the mutagenesis and enzymatic mapping data, these results indicate that the 2'OH on

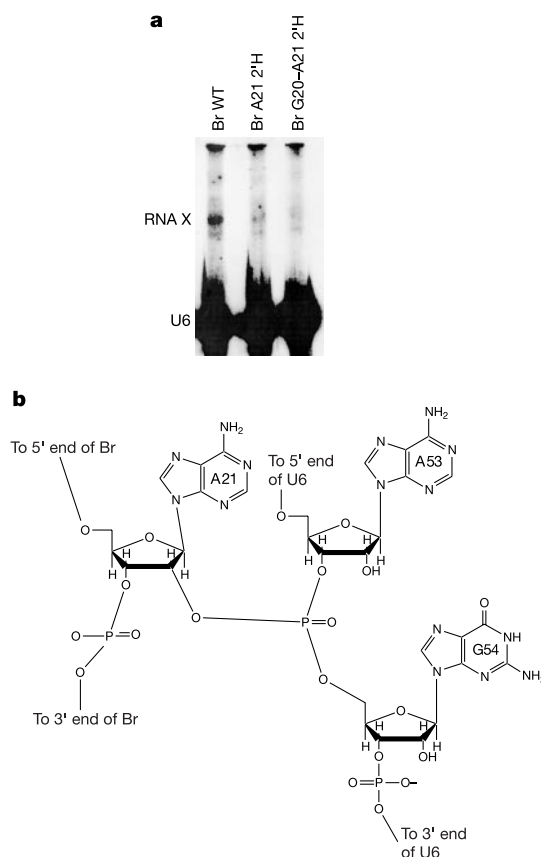


Figure 6 Requirement of the branchpoint 2'OH: a model for the linkage between U6 and Br in RNA X. **a**, Effect of 2'-deoxyribonucleoside substitutions in Br on RNA X formation. The identity of the Br derivative used in each reaction is indicated above each lane. The concentration of Br derivatives was 20 μ M and that of U2 and U6 was 0.1 μ M. WT, wild type. **b**, A model for the covalent bond between U6 and Br. The 2'OH of A21 in Br forms a triester with the phosphate between A53 and G54 in U6.

A21 participates directly in the covalent link between U6 and Br in RNA X (Fig. 6b).

Discussion

Our data show that a protein-free, base-paired RNA complex containing fragments of U2 and U6 snRNAs can catalyse a novel reaction similar to the first step of splicing, which leads to the formation of an X-shaped product (RNA X). Our results indicate that RNA X contains a 2'OH-phosphoester bond similar to that observed in lariat splicing products, but with the difference that in RNA X the 5' end of the branch acceptor (U6) is not displaced, thereby creating an unusual phosphotriester. The 2'OH of U6 A53 is close to the triester linkage, providing an explanation for the alkali sensitivity of RNA X (see Fig. 6b). A large, branched RNA species such as RNA X is probably stabilized by extensive stacking interactions³² and by cations that neutralize the negative charge repulsion in the compact branch core. It is plausible to think that a divalent cation, or another interaction, keeps the 2'OH of A53 protonated to prevent it from attacking the triester structure at neutral pH. The sensitivity of RNA X to complete digestion with RNases is consistent with a triester linkage, as the 2'OH on A53 can be activated to attack the triester, resulting in its dissociation to a diester, which can be further cleaved to nucleotides. The low efficiency of product formation can also be explained by the unusual chemistry, involving extraction of H₂O rather than the 5' end of U6 providing the leaving group, which would have created a Y-like structure with a 2' to 5' linkage. It is noteworthy that in ligase reactions the ultimate leaving group is H₂O, although the reaction involves multiple steps³³. Formation of RNA X may also involve more than one step. Lack of a 5' splice-site acceptor in our system may place the non-bridging phosphate oxygen of G54 in the pocket intended for the leaving group, thereby promoting formation of the triester rather than a transesterification reaction.

Our data demonstrate the ability of protein-free spliceosomal snRNAs to catalyse a reaction related to the first step of splicing, with the 2'OH of a bulged A residue in a base-paired, intron-like RNA attacking the phosphate 5' to a G residue in the invariant AGC triad of U6. The strong similarities in the sequence requirements between this reaction and the first step of splicing indicate similarities in the active site of the two reactions, with the ACAGAG and the AGC invariant regions of U6 located in close proximity to the reacting groups and having important roles in the chemistry of a metal-dependent reaction in both systems. Aberrant splicing reactions in the spliceosome in which the backbone of U6, rather than the 5' splice site, is attacked by the branch A are not unprecedented³⁴. Furthermore, the location of the covalent bond in U6 (3' of A53) is immediately adjacent to the site of an intron insertion in a fungal U6 gene, which is believed to have resulted from a faulty splicing reaction³⁵. In any event, our data point to the competence of the spliceosomal U2 and U6 snRNAs to function as the catalytic active site for the first step of splicing, and provide direct evidence for RNA-based catalysis in the spliceosome. □

Methods

Transcription and end labelling

U2 and U6 snRNA fragments were transcribed and the 5' ends were labelled as described²³. Labelling of the 3' end was done according to published procedures³⁶. Branch oligonucleotides were made by *in vitro* transcription (using a T7 transcript with extra nucleotides added at the 3' end for efficiency reasons, followed by DNase³⁷ removal of these extra nucleotides and gel purification) and by chemical synthesis (Dharmacon Biotech). We carried out end labelling as described for U2 and U6.

Catalytic assays

U2 and U6 were annealed as described²³ in a buffer containing 40 mM Tris, pH 7.2 and 40 mM MgCl₂, at a final concentration of 2 µM, followed by the addition of Br at a final concentration of 10 nM in a typical reaction. Reaction mixtures were routinely incubated at room temperature for 24 h and were analysed by 12% or 20% denaturing PAGE. For

large-scale reactions, the concentration of Br was increased to 9 µM, and after incubation at room temperature for 24 h, RNA X was purified from a 12% denaturing gel. Reactions for 3' labelling were as described with oligonucleotides specific to U6 or Br³⁶. The purified products were directly used in kinasing reactions for 5' labelling on Br. To obtain products that were 5'-labelled on U6, RNA X was dephosphorylated before the 5' labelling reaction, resulting in simultaneous labelling of the 5' ends of both U6 and Br, and repurified.

We carried out reverse transcription, RNase T1 and alkaline hydrolysis reactions as described²³. To hydrolyse the link between U6 and Br in RNA X, alkali hydrolysis reactions²³ were done at room temperature for 20 min. RNase T2 (Gibco) reactions were done in a buffer containing 10 mM sodium citrate pH 5.0, 0.5 mM EDTA and 1.5 M urea at 55 °C for 2 min.

Phosphorothioate substitution studies

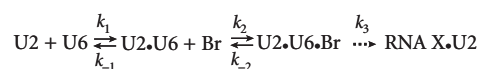
Transcription of U6 with α-thio-NTPs (Trilink Biotech and NEN) was done as described²⁷ so that each molecule would contain on average only a single thio group. Separate *in vitro* transcription reactions were set up for each α-S-NTP. The phosphorothioate-containing U6 was purified and treated with calf intestinal alkaline phosphatase to remove the 5' phosphate. Typical large-scale reactions were set up for each α-S-NTP-containing U6, and the resulting RNA X was gel-purified and labelled at the 5' end, resulting in labelling of both U6 and Br. Labelling of the 3' end of thio-U6-containing RNA X was done as described above, using a U6-specific oligonucleotide. Iodoethanol cleavage of the phosphorothioate-containing RNA X was done as described²⁷ and the reactions were loaded directly onto a 20% denaturing gel, along with identical reactions with similarly labelled U6 as a control.

Mutagenesis and binding studies

For mutagenesis studies, for each mutant the amount of RNA X formed was measured under standard reaction conditions in at least three independent experiments, and the average of normalized values was plotted as a fraction of wild type with two standard deviations as the error bars. For Br-binding studies, reactions were allowed to proceed for 2 h, after which they were divided in two; half was loaded on a denaturing 12% gel and the other half was loaded on a non-denaturing 8% PAGE run at 4 °C, followed by quantification of the reacted, bound and unbound Br. The bound and reacted fractions were normalized to the total amount of Br in each lane. As the bound Br on native gels also contained the reacted fraction, the amount of unreacted bound fraction was calculated by subtracting the reacted fraction from it. To test the ability of Br derivatives to competitively inhibit wild-type Br, 1–10 µM of the Br species to be tested was added to a typical reaction containing labelled wild-type Br and U6–U2 complex. The quantity of labelled RNA X formed was then compared with similar reactions containing the same concentration of unlabelled wild-type Br. All quantifications were done using a Molecular Dynamics PhosphorImager, with the amount of RNA X normalized to the amount of input in each reaction.

Kinetics

To determine the reaction rate (k_{obs}) the natural log of substrate remaining (1 – fraction RNA X/fraction RNA X at the end point of the reaction) was plotted against time under standard conditions, and the slope of the line was calculated to determine k_{obs} . For those Br mutants that showed significant increases in binding to the U2–U6 complex, the catalytic constant, k_{cat} , of the reaction was determined. The amount of RNA X was measured after 2 h of reaction with a U2–U6 complex concentration of 0.1 µM and different concentrations of the branch oligonucleotide (10 nM to 50 µM), and the apparent K_m and V_{max} (velocity of enzyme-catalysed reaction at infinite concentration of substrate) of the reactions were determined by plotting the data on an Eisenthal–Cornish-Bowden plot. The value for k_{cat} was calculated as $V_{\text{max}} / \text{the concentration of the U2–U6 complex, [U2U6]}$. The reaction was assumed to proceed according to the equation (dots represent non-covalent binding)



Received 19 June; accepted 10 August 2001.

- Nilsen, T. W. In *RNA Structure and Function* (eds Simons, R. & Grunberg-Manago, M.) 279–307 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1998).
- Burge, C. B., Tuschl, T. H. & Sharp, P. A. In *RNA World II* (eds Gesteland, R. E., Cech, T. R. & Atkins, J. E.) 525–560 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1998).
- Hetzer, M., Wurtzer, G., Schweyen, R. J. & Mueller, M. W. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature* **386**, 417–420 (1997).
- Collins, C. A. & Guthrie, C. The question remains: is the spliceosome a ribozyme? *Nature Struct. Biol.* **7**, 850–854 (2000).
- Newman, A. RNA splicing: out of the loop. *Curr. Biol.* **7**, R418–R420 (1997).
- Yean, S. L., Wuenschell, G., Termini, J. & Lin, R. J. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881–884 (2000).
- Sontheimer, E. J., Sun, S. & Piccirilli, J. A. Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801–805 (1997).
- Gordon, P. M., Sontheimer, E. J. & Piccirilli, J. A. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA* **6**, 199–205 (2000).
- O'Keefe, R. T., Norman, C. & Newman, A. J. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* **86**, 679–689 (1996).

10. Segault, V. *et al.* Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. *Mol. Cell. Biol.* **19**, 2782–2790 (1999).
11. Fabrizio, P. & Abelson, J. Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science* **250**, 404–409 (1990).
12. Madhani, H. D., Bordonne, R. & Guthrie, C. Multiple roles for U6 snRNA in the splicing pathway. *Genes Dev.* **4**, 2264–2277 (1990).
13. Datta, B. & Weiner, A. M. The phylogenetically invariant ACAGAGA and AGC sequences of U6 small nuclear RNA are more tolerant of mutation in human cells than in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 5377–5382 (1993).
14. Wolff, T., Menssen, R., Hammel, J. & Bindereif, A. Splicing function of mammalian U6 small nuclear RNA: conserved positions in central domain and helix I are essential during the first and second step of pre-mRNA splicing. *Proc. Natl Acad. Sci. USA* **91**, 903–907 (1994).
15. Sun, J. S. & Manley, J. L. A novel U2–U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9**, 843–854 (1995).
16. Fabrizio, P. & Abelson, J. Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing *in vitro*. *Nucleic Acids Res.* **20**, 3659–3664 (1992).
17. Yu, Y. T., Maroney, P. A., Darzynkiwicz, E. & Nilsen, T. W. U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *RNA* **1**, 46–54 (1995).
18. Collins, C. A. & Guthrie, C. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev.* **13**, 1970–1982 (1999).
19. Siatecka, M., Reyes, J. L. & Konarska, M. M. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983–1993 (1999).
20. Tuschl, T., Sharp, P. A. & Bartel, D. P. Selection *in vitro* of novel ribozymes from a partially randomized U2 and U6 snRNA library. *EMBO J.* **17**, 2637–2650 (1998).
21. Tuschl, T., Sharp, P. A. & Bartel, D. P. A ribozyme selected from variants of U6 snRNA promotes 2',5'-branch formation. *RNA* **7**, 29–43 (2001).
22. Madhani, H. D. & Guthrie, C. Randomization-selection analysis of snRNAs *in vivo*: evidence for a tertiary interaction in the spliceosome. *Genes Dev.* **8**, 1071–1086 (1994).
23. Valadkhan, S. & Manley, J. L. A tertiary interaction detected in a human U2–U6 snRNA complex assembled *in vitro* resembles a genetically proven interaction in yeast. *RNA* **6**, 206–219 (2000).
24. Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**, 1824–1832 (1996).
25. Tarn, W. Y. & Steitz, J. A. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**, 132–137 (1997).
26. Gish, G. & Eckstein, F. DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* **240**, 1520–1522 (1988).
27. Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091–7104 (1998).
28. Burn, A. J. & Cadogan, J. I. G. The reactivity of organophosphorus compounds. Part IX, the reaction of thionates with alkyl iodides. *J. Chem. Soc.* 5532–5541 (1961).
29. Saenger, W. *Principles of Nucleic Acid Structure* 116–126 (Springer, New York, 1984).
30. Query, C. C., Strobel, S. A. & Sharp, P. A. Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392–1402 (1996).
31. Query, C. C., Moore, M. J. & Sharp, P. A. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587–597 (1994).
32. Koole, L. H., Agback, P., Glemarec, C., Zhou, X.-X. & Chattopadhyaya, J. Solution structure of pentameric and heptameric branched-RNA modelling the lariat structure of group II or nuclear m-RNA introns studied by one- and two-dimensional NMR spectroscopy at 500 MHz. *Tetrahedron* **41**, 3183–3206 (1991).
33. Lehman, I. R. DNA ligase: structure, mechanism, and function. *Science* **186**, 790–797 (1974).
34. Yu, Y. T., Maroney, P. A. & Nilsen, T. W. Functional reconstitution of U6 snRNA in nematode *cis*- and *trans*-splicing: U6 can serve as both a branch acceptor and a 5' exon. *Cell* **75**, 1049–1059 (1993).
35. Tani, T. & Ohshima, Y. mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. *Genes Dev.* **5**, 1022–1031 (1991).
36. Huang, Z. & Szostak, J. W. A simple method for 3'-labeling of RNA. *Nucleic Acids Res.* **24**, 4360–4361 (1996).
37. Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA* **94**, 4262–4266 (1997).

Acknowledgements

We thank K. Ryan for advice and helpful discussions and insights, T. Nilsen for suggestions and comments on the manuscript, and A. Pyle for helpful discussions, advice and suggestions. We also thank G. Zubay, D. Herschlag, T. Tuschl, M. Konarska, G. Just, H. Lonnberg and F. Eckstein for discussions and advice on the chemistry of the reaction, and H. Robertson and D. Bartel for discussion and comments. We also thank P. Nouri for technical assistance. This work is supported by the NIH.

Correspondence and requests for materials should be addressed to J.L.M. (e-mail: jlm2@columbia.edu).

Discovery of X-rays from the protostellar outflow object HH2

Steven H. Pravdo^{*}, Eric D. Feigelson[†], Gordon Garmire[‡], Yoshitomo Maeda[†], Yohko Tsuboi[†] & John Bally[‡]

^{*} Jet Propulsion Laboratory, California Institute of Technology, Mail Stop 306-438, 4800 Oak Grove Drive, Pasadena, California 91109, USA

[†] 525 Davey Laboratory, Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, Pennsylvania 16802-6305, USA

[‡] Astrophysics and Planetary Science, University of Colorado, Boulder, Colorado 80309, USA

Herbig–Haro (HH) objects have been known^{1,2} for 50 years to be luminous condensations of gas in star-forming regions, but their underlying physical nature is still being elucidated. Previously suggested models encompass newborn stars³, stellar winds clashing with nebular material⁴, dense pockets of interstellar gas excited by shocks from outflows⁵, and interstellar ‘bullets’ (ref. 6). Recent progress has been made with the jet-induced shock model⁷, in which material streams out of young stellar objects and collides with the surrounding interstellar medium. A clear prediction of this model is that the most energetic Herbig–Haro objects will emit X-rays, although they have not hitherto been detected⁸. Here we report the discovery of X-ray emission from one of the brightest and closest Herbig–Haro objects, HH2, at a level consistent with the model predictions. We conclude that this Herbig–Haro object contains shock-heated material located at or near its leading edge with a temperature of about 10^6 K.

Herbig–Haro objects exhibit emission at many wavelengths that provide evidence for highly excited material. Optical emission from hydrogen Balmer lines, [O I], [S II], [N II], and [O III] traces shock-

heated plasma with temperatures ranging from several thousand to over one hundred thousand kelvin. Ultraviolet emission from C IV, C III], and N III] has been the highest excitation species detected⁹ (where a pair of square brackets indicates a forbidden transition and one indicates a semi-forbidden transition). Radio free-free emission is also detected¹⁰ in the most energetic objects. Most Herbig–Haro objects appear to be driven by highly collimated bipolar jets powered by accreting protostars and many exhibit large proper motions with velocities of hundreds of kilometres per second away from nearby young stellar objects. Although the temperatures of the post-shock plasmas in high-excitation Herbig–Haro objects are expected to reach values of the order of a million degrees, X-rays had not been seen because past instruments lacked the necessary spatial resolution and sensitivity.

We observed the region of Orion containing HH1 and HH2 for 21,400 s with the advanced charge-coupled device (CCD) imaging spectrometer (ACIS) of the Chandra X-ray observatory on 8 October 2000. We used the ACIS-S3 CCD for higher sensitivity to X-rays between 0.1–2 keV. Figure 1 illustrates the portion of the ACIS field of view containing HH2. There are several X-ray-emitting young stellar objects in this area of the Orion molecular cloud about 1 degree south of the Trapezium. The figure shows the Herbig Ae star V380 Ori, the classical T Tauri star known as the Cohen–Schwartz (CS) star, and other pre-main sequence stars newly identified in this observation.

X-ray sources on the ACIS image were identified in two bands, 0.1–2 keV and 2–8 keV, using a wavelet-based source detection algorithm¹². A source near the location of HH2 was identified in the soft band only. Despite the detection of only 11 events (photons) from this source within a 4-pixel (2 arcsec) radius, its reality is beyond question. In source-free regions of the image, the average count rate within the same radius is only 0.2, making the probability of detecting 11 photons from a blank field vanishingly small. The possibility that these events were caused by a flaring pixel or other known instrumental artefact was ruled out.

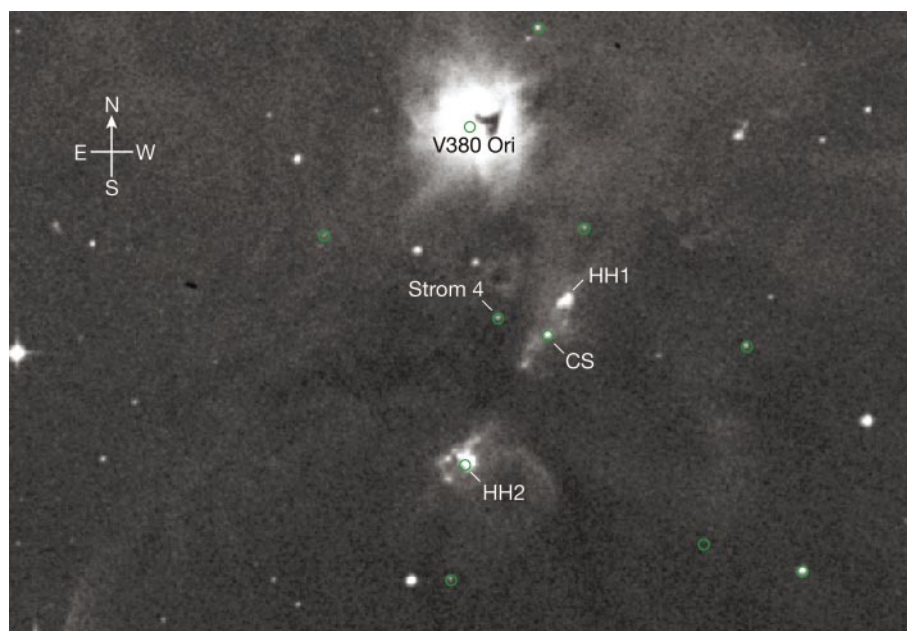


Figure 1 This figure shows the Orion Nebula region containing Herbig-Haro 2. The optical image is taken from the second digital sky survey and depicts the centre of the ACIS field observed with the Chandra X-ray Observatory. Circles indicate soft (0.1–2 keV) X-ray sources detected with ACIS in this observation. The brightest optical object is the reflection nebula NGC 1999, and the X-ray source centred therein is V380 Ori. HH1 (not detected in X-rays) is the bright nebulosity located southwest of NGC 1999 and HH2 is

located further south and east of HH1. The X-ray source just southeast of HH1 is the Cohen–Schwartz (CS) star. Strom 4 is east of the CS star. The newly discovered HH2H X-ray source is shown within the optical image of HH2. The circles are 4 arcseconds in diameter for clarity and to provide the scale, but the source locations are much better determined, as discussed in the text.

Three criteria establish the connection between this source and HH2. First, our astrometry places the ACIS source (designated CXO 053625.6 – 064716) within about 0.5 arcsec of one of the brightest optical knots with a very high proper motion in HH2. We used X-ray detections of stars in the field to correct for a small (<1 arcsec) offset in the Chandra positions. Four nearby stars were detected in the soft X-ray band: V380 Ori, which appears in the Hipparcos¹³, Tycho¹⁴, and USNO A2 (United States Naval Observa-

tory; ref. 15) catalogues, and three other USNO A2 stars. The offsets calculated from these fiducials (standards of reference) agree to within 0.2 arcsec. We used the average offset value to correct the X-ray positions. Figure 2a shows the position of the X-ray source (circle) superimposed on a Hubble space telescope (HST) H α image of HH2 (J. Bally *et al.*, manuscript in preparation). Also shown (squares) are the locations of 6-cm radio-continuum sources detected with the Very Large Array (VLA)¹⁶. Figure 2b shows the

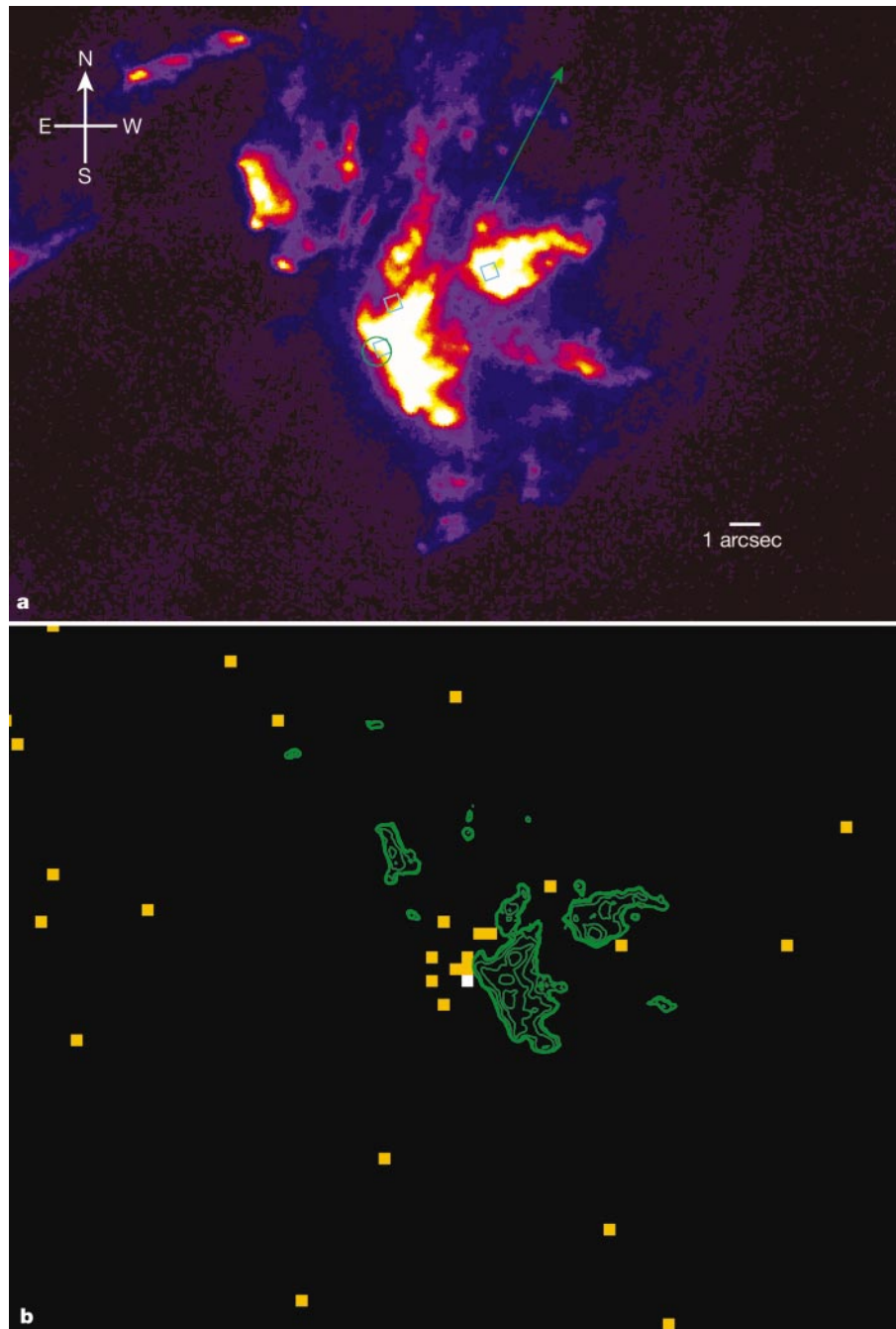


Figure 2 Expanded views of the region observed with the Hubble Space Telescope centred on Herbig–Haro 2. Northeast is to the upper left. **a**, The arrow pointing northwest indicates the direction to the HH2 energy source, the obscured young stellar object, VLA 1. The location of the ACIS source is shown as a circle (1 arcsec in diameter). The epoch of the ACIS observations is 2000. Squares indicate Very Large Array (VLA) sources whose original epoch was 1993, but are shown with their proper motions corrected to the ACIS epoch. The source VLA 10 coincides with the X-ray source. These source locations are

superimposed on the Hubble Space Telescope (HST) H α image of HH2 with epoch 1994. The proper motions of the HH2 H α knots are as large as $0.1 \text{ arcsec year}^{-1}$ so these knots moved as much as ~ 0.5 arcsec between the HST and the ACIS observations. **b**, This is an image of the ACIS HH2 X-ray source. It shows the individual X-ray events (orange squares, epoch 2000) from ACIS superimposed on the H α image contours (green lines) of HH2 observed with the HST (epoch 1994 and not proper-motion-corrected). The source VLA 10 (white square) is proper-motion-corrected to the ACIS epoch.

locations of individual X-ray photon events superimposed on a contour map of the H α emission. The X-rays coincide with the proper-motion-corrected positions of the radio source VLA 10 and the optical knot HH2H. The image in Fig. 2a shows the location of this H α knot in 1994 as observed using the HST. The X-ray and radio sources appear to be located near the leading edge of the optical knot.

Second, the X-rays are unlikely to originate from a point source. Figure 3 shows encircled energy plots for the HH2 X-ray source compared to two stars located nearby, the CS star and Strom 4 (see Fig. 1). A Kolmogorov–Smirnov test shows a 99% probability that the spatial distribution of the photons from HH2H is different from that of the stars. This indicates that these X-rays do not arise from a single point source and are consistent with an extended region such as HH2H.

Third, the spectrum of the HH2H source is unique in this field. Events are detected only below 2 keV. The observed X-ray spectrum can be fitted with a thermal plasma model (MEKA-L¹⁷) that has two free parameters—the plasma temperature and the column density of absorbing material. The best fit indicates $T = 1.2 \times 10^6$ K with an upper limit of 2.7×10^6 K. A plasma with a temperature of around 10^6 K is the minimum value to which ACIS is sensitive. The fitted column density has an upper limit of 9×10^{20} cm⁻² H atoms, equivalent to $A_v \approx 0.68$ magnitudes of visual extinction, implied by the reddening toward HH2H¹⁸. Such a thermal spectrum is expected from a high-temperature shock.

The positional coincidence of the X-ray source with the optical emission-line knot HH2H and the radio source VLA 10, the spatial extent of the X-ray emission, and the unique X-ray spectrum are convincing evidence that this X-ray source is associated with HH2H. This identification places the X-ray emission near the leading edge of the overall HH2 structure. However, its location with regard to the substructure of HH2H, that is, whether it is at the leading edge of HH2H, requires more sensitive and contemporaneous observations.

Why is HH2H the only location in the extensive HH1 and HH2 outflow to be detected in the X-ray spectrum? Four knots, three in HH2 (including HH2H) and one in HH1, are detected at 6 cm; these are the highest excitation knots in the HH1 and HH2 system. HH2H is distinguished from the other optically bright knots by its relatively strong ultraviolet emission¹⁹. Furthermore, HH2H exhibits among the largest proper motions in the system, 230 km s⁻¹ from optical measurements (430 km s⁻¹ from radio measurements), and it brightened optically relative to the other knots during the 1990s (ref. 20). Finally, there is evidence that the extinction towards

HH2H may be lower than towards the other high-excitation radio-bright knots. Thus, a combination of low extinction and high shock speeds may be why HH2H is seen in X-rays while the other knots are not. The upper limits for the other HH1 and HH2 knots are about 5 times lower than the HH2H detection.

At a distance of 480 pc, the X-ray luminosity of HH2H is $L_X = (0.41\text{--}5.2) \times 10^{29}$ erg s⁻¹ depending upon the amount of foreground absorption. The largest value of luminosity and the lowest value of temperature were obtained by accepting the measured upper limit to foreground absorption, which is consistent with the optical result. Therefore, we adopt $T = 10^6$ K and $L_X = 5.2 \times 10^{29}$ erg s⁻¹ for what follows.

In the jet-driven bow-shock model for Herbig–Haro objects, the leading edge of HH2H interacts with either stationary or slower-moving material in a bow shock. A relative shock velocity of $v_s \approx 200$ km s⁻¹ would result in an X-ray temperature of $T = 3\mu m_p v_s^2/16k \approx 1 \times 10^6$ K, as observed, where μ is the mean molecular mass and m_p is the proton mass. Thus, the detection of X-rays from this object implies that the shock is advancing into a medium that is slowly moving with respect to the ambient cloud. This velocity is the difference between the relative shock velocity and the proper motion.

The projected linear extent of the X-ray-emitting region is ~ 2 arcsec, or $r = 1.4 \times 10^{16}$ cm at 480 pc. Most of the X-rays are produced in the adiabatic part of the shock where the density can be estimated, in the simplest model, from the radiative cooling rate²¹, R , and the volume of the emitting regions, r^3 , as $n_X = (L_X/Rr^3)^{1/2} \approx 55$ cm⁻³. The radiative cooling time of the X-ray plasma is about equal to the time it takes material behind the shock to traverse HH2H, $t_c < r/V_s/4 \approx 90$ years. In 90 years the knot would move about 17 arcsec in the sky, giving observers ample opportunity to discover whether it remains integral (bullet-like) and cools, or whether it breaks apart and transmits its energy to other material. The total thermal energy content of the X-ray plasma behind the shock front is roughly $E < n_X k T r^3 = 2 \times 10^{40}$ ergs (about 0.4% of the kinetic energy associated with the mass responsible for the optical emission). The mass of hot plasma is around 1.2×10^{-7} solar masses, with the simple assumption of a uniform density.

The bow-shock model for the optical and ultraviolet lines in HH2H is already quite successful²², resulting in good agreement between measured and predicted intensities. That model has $v_o = 160$ km s⁻¹, $T_o = 1$ (H α) and 3 ([O III]) $\times 10^4$ K, and post-shock density, $n_o = 2,000$ cm⁻³. To first order, the cooling layer behind the shock is isobaric, so as the temperature decreases, the

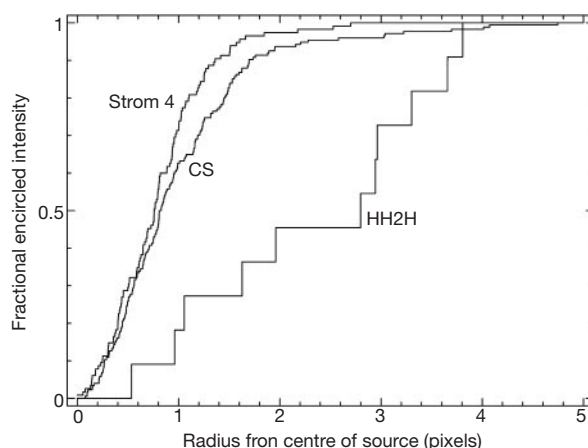


Figure 3 The HH2 X-rays are extended. The comparison of the spatial extent of HH2H with two nearby stars (Strom 4 and Cohen–Schwartz) in the 0.1–2 keV band shows that the

HH2 X-ray distribution is dissimilar at the 99% confidence level from those originating in point sources.

density increases. This is reasonable, since the cooling timescale of the X-ray gas is long compared with the dynamical timescale of the system (see above). Thus, from the X-ray derivation, the average density in the region producing the bulk of the optical emission lines should be $n_0 = n_X T/T_0 \approx 30\text{--}100 n_X$ or about $1,700\text{--}5,000 \text{ cm}^{-3}$. These values are consistent with the optical model. The difference between the shock velocities in the optical model and that derived for the X-rays could be real or could indicate a somewhat lower X-ray temperature, $T \approx 0.6 \times 10^6 \text{ K}$. A consistent theory of the emission from HH2H from optical through X-rays appears within reach.

However, it is also possible that while the X-ray emission arises from shock-heated post-shock fluid, the optical emission is dominated by a reverse shock propagating back into the fast material entering the HH2H region from the source (upstream) direction. On a much larger and faster scale, reverse shocks²³ are successful in explaining emission from supernova remnants. If the fast-moving upstream fluid is much denser than the downstream fluid before entering the reverse shock, the reverse shock speed will be much lower than the forward shock speed, and the post-shock material will predominantly radiate in lower excitation ultraviolet, optical and infrared lines. Deeper X-ray observations and nearly simultaneous optical images are needed to determine if the hot and relatively tenuous X-ray plasma does indeed lie at the leading (downstream) edge of the HH2H shock. □

Received 11 May; accepted 5 September 2001.

- Herbig, G. H. The spectra of two nebulous objects near NGC 1999. *Astrophys. J.* **113**, 697–698 (1951).
- Haro, G. Herbig's nebulous objects near NGC 1999. *Astrophys. J.* **115**, 572–573 (1952).
- Ambarzumian, V. A. *Commun. Burakan Obs.* **13**, 3–35 (1954).
- Canto, J. & Rodriguez, L. A stellar wind focusing mechanism as an explanation for Herbig–Haro objects. *Astrophys. J.* **239**, 982–987 (1980).
- Schwartz, R. D. A shocked cloudlet model for Herbig–Haro objects. *Astrophys. J.* **233**, 884–900 (1978).
- Norman, C. & Silk, J. Interstellar bullets: H₂O masers and Herbig–Haro objects. *Astrophys. J.* **228**, 197–205 (1979).
- Reipurth, B. & Bertout, C. (eds) *Herbig–Haro Flows and the Birth of Stars* (IAU Symp. 182, Kluwer, Chamonix, France, 1997).
- Pravdo, S. H. & Marshall, F. E. An X-ray active region in Orion: X-rays from a Herbig–Haro object? *Astrophys. J.* **248**, 591–595 (1981).
- Ortali, S. & Ordorico, S. A discussion on the nature of the Herbig Haro object no. 1 from its far UV spectrum. *Astron. Astrophys.* **83**, L8–L9 (1980).
- Pravdo, S. H. *et al.* Detection of radio continuum emission from the Herbig–Haro objects 1 and 2 and from their central exciting source. *Astrophys. J.* **293**, L35–L38 (1985).
- Bautz, M. W. *et al.* X-ray CCD calibration for the AXAF CCD imaging spectrometer. *Proc. SPIE* **3444**, 210–224 (1998).
- Chandra Interactive Analysis of Observations (CIAO), <http://asc.harvard.edu/ciao/>, CIAO 2 Science Threads, Detecting Sources in Imaging Observations.
- The Hipparcos and Tycho Catalogs Vols 1–17* (ESA SP-1200, European Space Agency, 1997).
- Hog, E. *et al.* The Tycho-2 catalogue of the 2.5 million brightest stars. *Astron. Astrophys.* **35**, L27–L30 (2000).
- Monet, D. G. The 526,280,881 objects in the USNO-A2.0 catalog *Am. Astron. Soc.* **193**, 12003 (1998).
- Rodriguez, L. F. *et al.* New VLA observations of the HH 1–2 region: evidence for density enhancements moving along the axis of the VLA 1 radio jet. *Astrophys. J.* **119**, 882–889 (2000).
- Mewe, R., Gronenschild, E. B. H. M. & van den Oord, G. H. J. Calculated X-radiation from optically thin plasmas. *V. Astron. Astrophys.* **62**, 197–254 (1985).
- Böhm, K. H. & Brugel, E. W. The iron abundance in Herbig–Haro objects and some new spectrophotometric data on H-H 2H. *Astron. Astrophys.* **74**, 297–301 (1979).
- Böhm-Vitense, E., Böhm, K. H., Cardelli, J. A. & Nemec, J. M. The ultraviolet and continuous emission of the Herbig–Haro objects HH 2 and HH 1. *Astrophys. J.* **262**, 224–233 (1982).
- Eisloffel, J., Mundt, R. & Böhm, K.-H. Structure and proper motions in Herbig–Haro objects 1 and 2. *Astron. J.* **108**, 1042–1055 (1994).
- Cox, D. P. & Daltabuit, E. Radiative cooling of a low-density plasma. *Astrophys. J.* **167**, 113–117 (1971).
- Hartigan, P., Raymond, J. & Hartmann, L. Radiative bow shock models of Herbig Haro objects. *Astrophys. J.* **316**, 323–348 (1987).
- McKee, C. F. X-ray emission from an inward-propagating shock in young supernova remnants. *Astrophys. J.* **188**, 335–339 (1974).

Acknowledgements

We thank L. Townsley and P. Broos for their assistance. We also thank B. Reipurth and R. Blandford for helpful discussions. This research was performed in part by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Correspondence and requests for materials should be addressed to S.P. (e-mail: spravdo@jpl.nasa.gov).

Suppression of crystal nucleation in polydisperse colloids due to increase of the surface free energy

Stefan Auer & Daan Frenkel

FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands

The formation of small crystallites is governed by two competing factors: the free energy gained upon transferring constituent atoms, molecules or colloidal particles from the metastable liquid to the more stable solid, and the free energy needed to create the surface area of the crystallite¹. Because the ratio of surface area to bulk is large for small particles, small crystallites dissolve spontaneously under conditions where larger crystallites are stable and macroscopic crystal growth occurs only if spontaneously formed crystallites exceed a critical minimum size. On theoretical grounds¹, the probability of forming such critical crystal nuclei is expected to increase rapidly with supersaturation. However, experiments show^{1,2} that the rate of crystal nucleation in many systems goes through a maximum as the supersaturation is increased. It is commonly assumed that the nucleation rate peaks because, even though the probability of forming critical nuclei increases with increasing concentration, the rate of growth of such nuclei decreases. Here we report simulations of crystal nucleation in suspensions of colloidal spheres with varying size distributions that show that the probability that critical nuclei will form itself goes through a maximum as the supersaturation is increased. We find that this effect, which is strongest for systems with the broadest particle size distribution, results from an increase with supersaturation of the solid–liquid interfacial free energy. The magnitude of this effect suggests that vitrification at high supersaturations should yield colloidal glasses that are truly amorphous, rather than nano-crystalline.

Colloidal suspensions of identical, hard, spherical particles can be either fluid or crystalline. At low densities, the fluid state is stable, but when the colloids occupy more than 49.4% of the volume, a crystalline phase should form². In practice, several factors influence the crystallization of hard-sphere colloids. First of all, synthetic colloids have a distribution of particle radii with a width that is rarely less than 2–3% of the average radius. This non-uniformity of size (“polydispersity”) is known to affect the location of the freezing curve. Simulations³ show that higher compressions are needed to freeze a polydisperse suspension. Irrespective of the composition of the coexisting fluid, the polydispersity of the crystal never exceeds 5.7%. Experiments on crystal formation in hard-sphere colloids indicate that crystallization is suppressed in suspensions with a polydispersity exceeding 12% (ref. 2). This must be due to kinetic factors, as crystallization of strongly polydisperse suspensions is not excluded on thermodynamic grounds.

Classical nucleation theory (CNT)¹ offers a simple thermodynamic explanation for why small crystal nuclei are less stable (that is, they have a higher free energy) than the supersaturated parent phase. CNT uses macroscopic arguments to estimate the free energy required to form a crystallite. The decrease in free energy due to the transfer of N particles from the metastable liquid to the solid state is approximated as $N\Delta\mu$, where $\Delta\mu = \mu_{\text{solid}} - \mu_{\text{liquid}}$ is the difference in chemical potential between the solid and the liquid state. The CNT estimate for the free-energy cost involved in the creation of the surface area A of the nucleus is γA , where γ is the surface free energy of the solid–liquid interface.

Owing to the competition between bulk and surface terms, the Gibbs free energy $\Delta G(N)$ required to form an N -particle nucleus

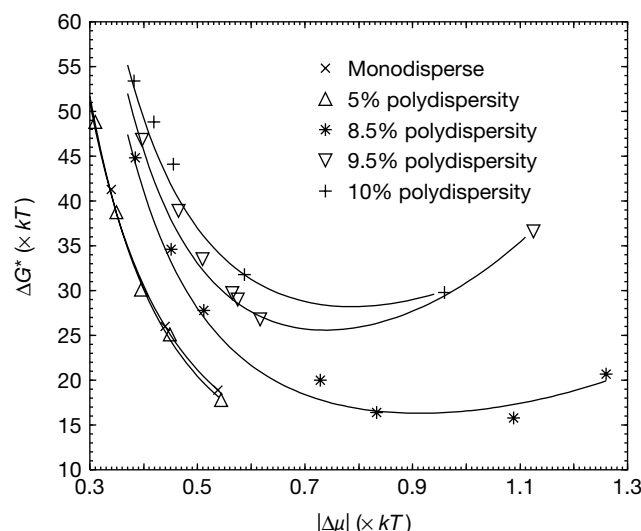


Figure 1 Computed free-energy barrier for crystal nucleation of polydisperse suspensions of hard, colloidal spheres. The free energy ΔG^* is expressed in terms of kT , where k is Boltzmann's constant and T is the absolute temperature. $|\Delta\mu|$ (also in terms of kT) is the absolute difference between the chemical potential of the liquid and the solid. It is a measure for the degree of supersaturation. The curves are fits that have been drawn as a guide to the eye. To facilitate comparison with experiment, we show, in Table 1, the relationship between $|\Delta\mu|$ and the volume fraction φ of the liquid for the different systems that we studied.

goes through a maximum at a value of N called the critical nucleus size. For a spherical nucleus, the maximum value of $\Delta G(N)$ is:

$$\Delta G^* = (16\pi/3)\gamma^3/(\rho|\Delta\mu|)^2 \quad (1)$$

where ρ is the number density of the crystal phase. The rate I at which nuclei are formed depends exponentially on ΔG^* :

$$I = \kappa \exp(-\Delta G^*/kT) \quad (2)$$

where T is the absolute temperature, k is Boltzmann's constant and κ is a kinetic prefactor that is proportional to the short-time self-diffusion constant of the colloids. The form of equation (2) does not rely on the validity of CNT.

In the CNT picture, increasing the supersaturation (that is, increasing $|\Delta\mu|$) lowers the nucleation barrier. If γ were independent of $|\Delta\mu|$, then ΔG^* would always decrease with increasing supersaturation. In experiments^{4–6} the rate of colloidal crystal nucleation starts to decrease again for large supersaturations. This effect is attributed to the decrease in the kinetic prefactor κ : in order to crystallize, colloidal fluids must be compressed beyond the freezing curve. But eventually, the suspension will vitrify under compression. This vitrification slows down the particle motion and

Table 1 Supersaturation and volume fraction of polydisperse colloids

0%		5%		8.5%		9.5%		10%	
$\Delta\mu$	φ	$\Delta\mu$	φ	$\Delta\mu$	φ	$\Delta\mu$	φ	$\Delta\mu$	φ
0.339	0.5207	0.310	0.5344	0.385	0.5614	0.397	0.5697	0.382	0.5717
0.439	0.5277	0.349	0.5377	0.451	0.5673	0.465	0.5746	0.419	0.5738
0.538	0.5342	0.395	0.5414	0.512	0.5726	0.509	0.5782	0.455	0.5775
		0.448	0.5456	0.728	0.5864	0.565	0.5808	0.587	0.5878
		0.544	0.5528	0.833	0.5948	0.575	0.5828	0.959	0.6239
				1.088	0.6145	0.616	0.5859		
				1.260	0.6212	1.125	0.6239		

$\Delta\mu$ is the supersaturation and φ is the volume fraction of the colloidal fluid. Pairs of values are given for polydispersities in the range from 0% (left) to 10% (right). The polydispersities quoted in Table 1 and in Figs 1 and 2 are those of the metastable liquid.

presumably reduces κ in equation (2). A problem with this interpretation is that recent experiments on colloidal crystallization in micro-gravity have found evidence for crystallization at densities that are well beyond the glass transition point⁷.

We performed Monte Carlo simulations to study the crystal-nucleation barrier and the structure of the critical nucleus, as a function of both polydispersity and supersaturation. As in the case of monodisperse suspensions⁸, we find that all critical nuclei have a randomly stacked close-packed structure. During crystallization, size-fractionation occurs³: the particles that make up the critical nucleus are on average larger than those in the metastable liquid. At fixed $|\Delta\mu|$ we find that ΔG^* , the height of the nucleation barrier, does not depend on the polydispersity for polydispersities $\leq 5\%$ (see Fig. 1). As the polydispersity is increased beyond 5%, ΔG^* increases rapidly. This implies that the probability of a critical nucleus forming is suppressed in polydisperse suspensions. It follows from equation (1) (or actually, from its polydisperse equivalent) that, at constant $|\Delta\mu|$, the variation of ΔG^* with polydispersity is due to an increase of the interfacial free energy γ . The increase of γ with polydispersity runs counter to Turnbull's suggestion that the interfacial free energy should be proportional to ΔH , the latent heat of fusion¹. For the systems that we studied, ΔH crosses zero at a polydispersity of about 9%, where the liquid becomes denser than the coexisting solid³. Yet, γ clearly remains non-zero.

Surprisingly, the variation of ΔG^* with $|\Delta\mu|$ is non-monotonic. As $|\Delta\mu|$ is increased, the nucleation barrier goes through a minimum (Fig. 1). This non-monotonic behaviour of ΔG^* is due to the increase of γ with $|\Delta\mu|$ (Fig. 2). To illustrate this, let us approximate the $|\Delta\mu|$ -dependence of γ by $\gamma \approx \gamma_0(1 + \alpha|\Delta\mu|)$. If we disregard the slight $|\Delta\mu|$ -dependence of the solid density, it then follows from equation (1) that ΔG^* must go through a minimum when $|\Delta\mu| = 2/\alpha$. The nucleation theorem⁹ suggests that the minimum in ΔG^* is due to the inversion of the number densities of the polydisperse fluid and the crystal nucleus. In CNT it is usually assumed that γ is constant. A linear variation of γ with $|\Delta\mu|$ has been observed in inorganic glasses¹, but there the constant α is negative and hence there is no minimum in ΔG^* . In other systems^{10,11}, non-monotonic behaviour of ΔG^* is induced by a hidden phase transition in the metastable phase.

The minimum value of ΔG^* increases rapidly with polydispersity. Using kinetic Monte Carlo simulations, we can estimate the value of

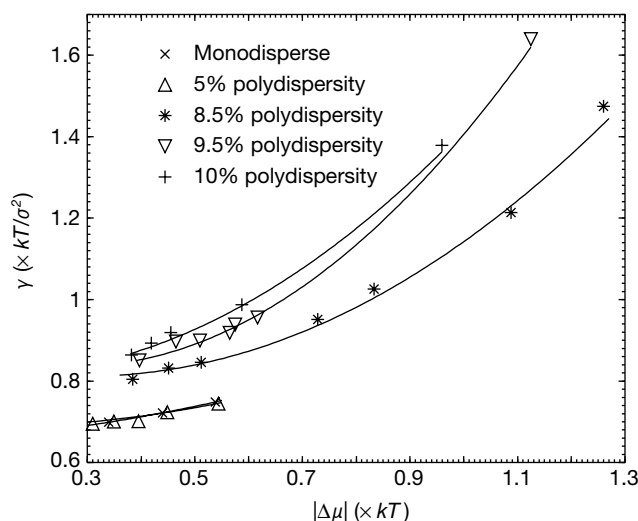


Figure 2 The interfacial free energy of crystal nuclei in polydisperse suspensions of hard, colloidal spheres. The interfacial free energy γ is expressed in terms of kT/σ^2 , where σ is the average hard-sphere diameter. The curves are fits that have been drawn as a guide to the eye.

the kinetic prefactor⁸. We find that, over the range of supersaturations studied, the kinetic prefactors vary by at most an order of magnitude (S.A. and D.F., manuscript in preparation). This means that the variation in the rate of nucleation is dominated by the behaviour of ΔG^* . We estimate that, for colloidal particles with a radius ≥ 500 nm, homogeneous nucleation will be effectively suppressed (less than one nucleus per cubic centimetre per day) when the polydispersity exceeds 10%. This finding has important implications for the morphology of polycrystalline colloidal materials. Using a simplified version of the analysis proposed in ref. 10 to estimate the size of crystallites in a polycrystalline sample, it is easy to derive that R_c , the average crystallite size at the end of a nucleation experiment, should scale as $\exp(\Delta G^*/4kT)$. Our observation of a minimum in ΔG^* thus implies the existence of a minimum in the typical crystallite size. This should be experimentally observable.

We could only compute ΔG^* if spontaneous nucleation did not occur in the course of a simulation. In practice, this implied that we could not study barriers lower than $15kT$. As a result, we could not test whether ΔG^* in systems with a low polydispersity (less than 8.5%) also has a minimum. If we assume that at lower polydispersities we can extrapolate the increase of γ with $|\Delta\mu|$ to large supersaturations, then we predict that a minimum in ΔG^* should occur even in nearly monodisperse systems. Again, this should be experimentally observable, because we should expect to see the formation of larger crystallites if the solution can be compressed rapidly through the region where ΔG^* is small.

There are two ways to interpret the experimental finding that crystallization is not observed in suspensions with a polydispersity greater than 12%: either crystals do not form, or they are too small to be observed. Our simulations support the first interpretation. In a suspension of colloids with a 500-nm radius, we would expect¹⁰ to see less than one crystallite per cubic centimetre, once $\Delta G^* > 32kT$. In other words, under those conditions the colloidal glass is truly amorphous.

Our predictions concerning the structure and free energy of colloidal crystal nuclei can be tested experimentally. Recently, the technique of confocal scanning laser microscopy has been used¹² to study the structure and size of critical crystal nuclei in dense colloidal suspensions and would thus be perfectly suited to test our predictions. Our prediction concerning the minimum in ΔG^* is even easier to verify. By visual inspection, one could verify whether the crystallites that nucleate in strongly supersaturated solutions are larger than those that form at lower supersaturations. Over a decade ago, Pusey and van Meegen produced images of the morphology of polycrystalline hard-sphere colloids¹³ and similar morphologies have recently been observed in a study of colloidal crystallization in micro-gravity (Z. D. Cheng, W. B. Russel and P. M. Chaikin, unpublished data). The observed increase of the crystallite size at large supersaturations was attributed to heterogeneous nucleation¹³, so a direct test of our prediction is still lacking. □

Methods

The simulation techniques that are required to compute the free energy of small crystal nuclei have been described in earlier publications^{8,14}. In the present work, we used constant-pressure, semi-grand-canonical Monte Carlo (SGMC) simulations of the type described in ref. 3. To eliminate possible finite-size effects, we used systems of 3,375 particles. Very long runs (up to 1.6×10^7 trial moves per particle), in combination with parallel tempering⁸, were needed to ensure equilibration of the dense, polydisperse fluid.

Received 16 July; accepted 3 September 2001.

1. Kelton, K. F. *Solid State Physics* Vol. 45 (eds Ehrenreich, H. & Turnbull, D.) 75–178 (Academic, New York, 1991).
2. Pusey, P. in *Liquids, Freezing and Glass Transition* (eds Hansen, J. P., Levesque, D. & Zinn-Justin, J.) 763–931 (North-Holland, Amsterdam, 1991).
3. Kofke, D. A. & Bolhuis, P. G. Freezing of polydisperse hard spheres. *Phys. Rev. E* **59**, 618–622 (1999).
4. Palberg, T. Crystallization kinetics of repulsive colloidal spheres. *J. Phys. Condens. Matter* **11**, 323–360 (1999).
5. Van Duijnvelde, J. S. & Lekkerkerker, H. N. W. in *Science and Technology of Crystal Growth* (eds van der Eerden, J. P. & Bruinsma, O. S. L.) 279–290 (Kluwer Academic, Dordrecht, 1995).
6. Harland, J. L. & Van Meegen, W. Crystallization kinetics of suspensions of hard colloidal spheres. *Phys. Rev. E* **55**, 3054–3067 (1997).

7. Zhu, J. et al. Crystallization of hard-sphere colloids under microgravity. *Nature* **387**, 883–885 (1997).
8. Auer, S. & Frenkel, D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature* **409**, 1020–1023 (2001).
9. Oxtoby, D. W. & Kashchiev, D. A general relation between the nucleation work and the size of the nucleus in multicomponent nucleation. *J. Chem. Phys.* **100**, 7665–7671 (1994).
10. Shi, F. G., Tong, H. Y. & Ayers, J. D. Free energy barrier to nucleation of amorphous-to-crystalline transformation selects the scale of microstructure of crystallized materials. *Appl. Phys. Lett.* **67**(3), 350–352 (1995).
11. Ten Wolde, P. R. & Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **277**, 1975–1978 (1997).
12. Gasser, U., Weeks, E. R., Schofield, A., Pusey, P. N. & Weitz, D. A. Real-space imaging of nucleation and growth in colloidal crystallization. *Science* **292**, 258–262 (2001).
13. Pusey, P. N. & Van Meegen, W. Phase behaviour of concentrated suspensions of nearly hard colloidal spheres. *Nature* **320**, 340–342 (1986).
14. Ten Wolde, P. R., Ruiz-Montero, M. J. & Frenkel, D. Numerical evidence for bcc ordering at the surface of a critical fcc nucleus. *Phys. Rev. Lett.* **75**, 2714–2717 (1995).

Acknowledgements

We thank A. van Blaaderen, P. M. Chaikin, H. N. W. Lekkerkerker, D. Oxtoby, W. B. Russel and F. Shi for comments and suggestions. This work was supported by the division of Chemical Sciences of the Netherlands Organization for Scientific Research (NWO). The work of the FOM Institute is part of the research programme of FOM and is made possible by financial support from the NWO. Access to the TERAS supercomputer was made possible through a grant by the NCF.

Correspondence and requests for materials should be addressed to D.F. (e-mail: Frenkel@amolf.nl).

Self-assembled monolayer organic field-effect transistors

Jan Hendrik Schön, Hong Meng & Zhenan Bao

Bell Laboratories, Lucent Technologies, Mountain Avenue, Murray Hill, New Jersey 07974, USA

The use of individual molecules as functional electronic devices was proposed in 1974 (ref. 1). Since then, advances in the field of nanotechnology have led to the fabrication of various molecule devices and devices based on monolayer arrays of molecules^{2–11}. Single molecule devices are expected to have interesting electronic properties, but devices based on an array of molecules are easier to fabricate and could potentially be more reliable. However, most of the previous work on array-based devices focused on two-terminal structures: demonstrating, for example, negative differential resistance⁸, rectifiers⁹, and re-configurable switching^{10,11}. It has also been proposed that diode switches containing only a few two-terminal molecules could be used to implement simple molecular electronic computer logic circuits¹². However, three-terminal devices, that is, transistors, could offer several advantages for logic operations compared to two-terminal switches, the most important of which is ‘gain’—the ability to modulate the conductance. Here, we demonstrate gain for electronic transport perpendicular to a single molecular layer (~ 10 – 20 Å) by using a third gate electrode. Our experiments with field-effect transistors based on self-assembled monolayers demonstrate conductance modulation of more than five orders of magnitude. In addition, inverter circuits have been prepared that show a gain as high as six. The fabrication of monolayer transistors and inverters might represent an important step towards molecular-scale electronics.

It has been shown that conjugated organic molecular materials behave very much like conventional semiconductors and various electronic devices have been demonstrated^{13,14}. Although great progress has been made, the knowledge of the conduction mechanism on the molecular scale is not completely understood yet⁵. Nevertheless, the possibility of three-terminal devices using a benzene ring and other molecules has been proposed^{15,16} and an

electromechanical amplifier using a single molecule has been demonstrated¹⁷. In this study, self-assembled monolayers (SAMs) of six conjugated molecules (Fig. 1a; molecules **1** to **6**) were used as the active material in field-effect transistors (FETs). The schematic device structure is shown in Fig. 1b. A highly doped Si substrate was used as the gate electrode. The substrate was structured using conventional lithography and anisotropic etching in order to define vertical steps¹⁸. The possible benefits of building vertical FETs on sidewalls of trenches or Si pillars have been recognized for at least a quarter century^{19,20}. A 30-nm layer of thermally grown SiO₂ is used as a gate insulator. In order to fabricate a source electrode a gold layer was deposited by thermal evaporation. SAMs of thiol- or isocyanide-terminated conjugated molecules were then formed on top of the gold film^{21,22}. Ellipsometry was used to confirm that only a monolayer has been deposited on top of the gold surface. The chemical structures of the self-assembled molecules are shown in Fig. 1a. The channel length is defined by the thickness of the SAM, similar to short-length vertical Si-FETs (ref. 20). Finally, the drain electrode was prepared by shallow-angle shadow evaporation of gold using a collimated beam and a vertical trench in the Si-substrate as shadow mask onto a cooled substrate (~100 K) to avoid damage to the SAM and form an active contact area of approximately $0.8 \times 0.1 \mu\text{m}^2$. Because the conduction in organic FETs takes place only within 50 Å from the semiconductor-insulator interface²¹, only several thousand molecules will be probed in such self-assembled monolayer field-effect transistors (SAMFETs), where the channel length is defined by the thickness of the monolayer. Devices based on all six compounds exhibited very

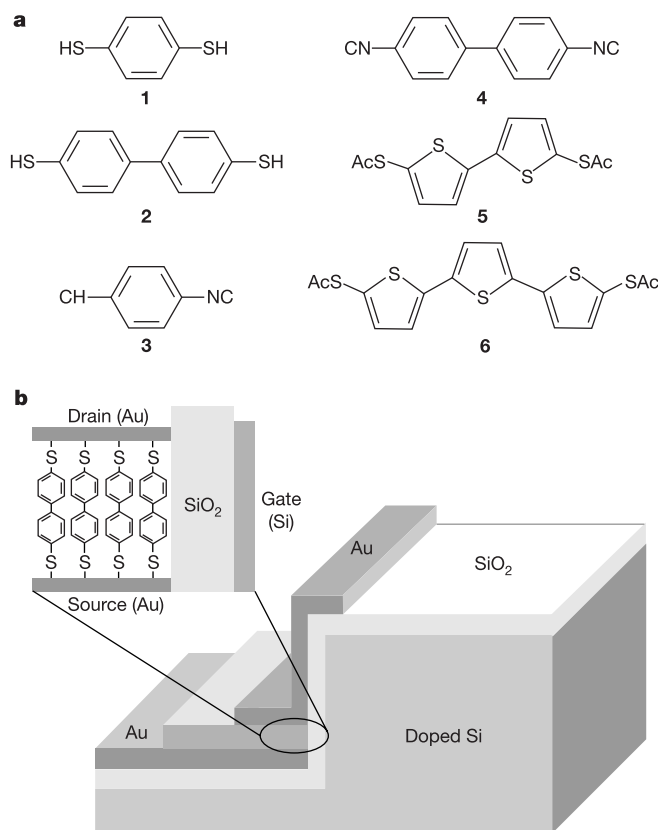


Figure 1 Structure of the investigated molecules and transistors. **a**, Molecular structure of the investigated materials; **b**, SAMFET structure: a highly doped Si-substrate is used as the gate electrode, a thermally grown SiO₂ layer acts as gate insulator, the gold source electrode is deposited by thermal evaporation, the active semiconducting material is a self-assembled monolayer (SAM) of one of the six molecules (**1**–**6**), and the drain contact is defined by shallow-angle shadow evaporation of gold. The active region of the device is magnified.

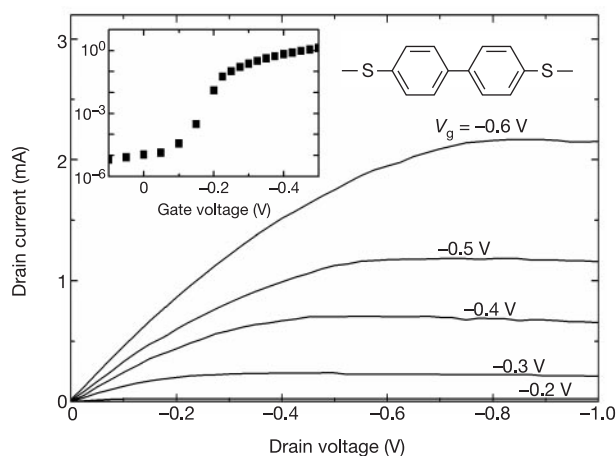


Figure 2 Transistor characteristics of a 4,4'-biphenyldithiol (molecule **2**) SAMFET at room temperature. The inset shows the transfer characteristics, that is, drain current at V_d = -1 V as a function of V_g.

similar performance and current–voltage characteristics. Here 4,4'-biphenyldithiol (molecule **1**) will be used as the model material and the device operation at room temperature is discussed.

Figure 2 shows typical transistor current–voltage characteristics of a 4,4'-biphenyldithiol SAMFET at room temperature. Similar behaviour has been observed for many devices giving statistical relevance to these data. The drain current (I_d) can be modulated by approximately five orders of magnitude by the applied gate bias (V_g). The device acts as a normally 'off' p-channel transistor, that is, a negative V_g increases the conductance of the channel, with a threshold voltage of approximately -0.2 V. No degradation effects have been observed up to 10⁵ cycles. The modulation ratio seems to be limited by the low intrinsic conductivity of the SAM (see inset to Fig. 2). The gate potential varies only the charge transport in molecules assembled close to the gate insulator (several thousand molecules, about 1% of the molecules in the monolayer)²³; this has been confirmed for devices with different contact areas and channel widths. However, the 'off' current samples all of the molecules covering the entire contact area (several hundred thousand molecules). Consequently, the experimentally observed on/off current ratio of 10⁵ of the whole device corresponds to a conductance change through an individual molecule by a factor of approximately 10⁷ (10⁵ × (100,000/1,000) = 10⁷) owing to the applied gate bias. A

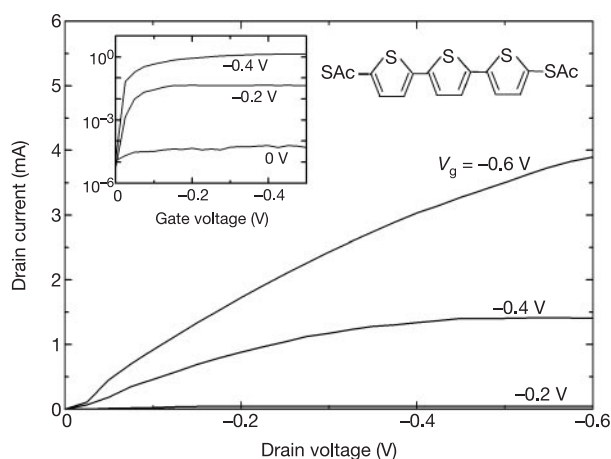


Figure 3 Transistor characteristics of molecule **6** SAMFET at room temperature. The inset shows the characteristics on a logarithmic scale. For V_g = 0 V, more or less linear and symmetric I–V characteristics are observed.

Table 1 Room temperature performance of several SAMFETs

Molecule	Transconductance, g (mA V^{-1})	Threshold voltage, V_t (V)	On/off current ratio
1	65	-0.25	10^6
2	130	-0.2	10^6
3	60	-0.25	10^5
4	120	-0.2	10^5
5	135	-0.25	10^5
6	170	-0.2	10^5

The SAMFETs are based on various SAMs: molecules **1** to **6**.

conductance of approximately $5 \mu\text{S}$ per molecule can be estimated, which is in reasonable agreement with calculations presented in the literature¹⁵. In addition, no resonant tunnelling features or charging effects were observed in the room-temperature characteristics of these SAMFETs. Several control experiments were performed using, for example, a highly insulating self-assembled alkanethiol monolayer, and no gate modulation of the conductance could be observed. Phenyl-based monothiolates resulted in diode-like current-voltage characteristics. Nonetheless, a significant gate modulation was also observed in these materials. It is difficult to explain the strong conductance modulation solely on the basis of quantum-mechanical treatment of the conduction through a monolayer^{15,24}. First-principles calculations indicate a modulation of more than one order of magnitude for p-phenylenedithiol (**2**)¹⁵. Additional effects could include a gate-modulation of the potential distribution along the channel length²⁵ or a suppression of the barrier at the contacts^{3,26}. Barrier heights within the range from 0.15 to 0.4 eV have been reported for isocyanide-metal junctions²⁷.

A transconductance (g) in the range of 120 to 130 mA V^{-1} can be calculated from the device characteristics of biphenyl-based SAMFETs at room temperature. Slightly lower values ($g \approx 60 \text{ mA V}^{-1}$) have been achieved for the monophenyls. Thiophene-based SAMFETs revealed even higher transconductances (see Fig. 3 and Table 1). No significant differences were observed between dithiols and diisocyanides indicating that the current modulation is related to the conjugated 'backbone' of the molecule. Nevertheless, these experimental values for the channel conductance and transconductance, taking the geometry factors into account, are significantly higher than those achieved for oligophenylene single crystals or related conjugated hydrocarbon systems^{28–30}. This indicates that the charge transport through the single monolayer of molecules determines the characteristics of SAMFETs rather than the transverse conduction from molecule to molecule as in the case of bulk organic crystals or thin films. Hence, one might envision much higher speeds with SAMFET-based circuits than in 'macroscopic' organic

electronic circuits. However, effects of parasitic capacitances have to be taken into account, which might limit device performance.

In addition, inverters have been fabricated using two 4,4'-biphenyldithiol SAMFETs (Fig. 3). Figure 4 shows the static switching characteristics of such an inverter circuit, which is a basic building block for more complex logic circuits. Here, the output switches from logic '1' (-2 V) to logic '0' (0 V) when the input switches from logic '0' to logic '1'. Moreover, a gain of six has been obtained from the characteristics of such an inverter circuit at room temperature.

The demonstration of gain in ultrashort field-effect transistors based on molecular monolayers can be seen as an important step towards molecular electronics. It opens up new opportunities for integrated molecular material electronic devices. We have also shown here that we can take advantage of conventional Si technologies in combination with molecular self-assembly techniques in order to realize new devices. The device structure used here can be applied to many other molecular systems, where ultra-small channel lengths can easily be defined by the thickness of the active semiconducting layer, which will allow the studies of transport in the nanometre scale. □

Received 4 May; accepted 17 August 2001.

- Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
- Goldhaber-Gordon, D., Montemero, M. S., Love, C. J., Opiteck, G. J. & Ellenbogen, J. C. Overview of nanoelectronic devices. *Proc. IEEE* **85**, 521–540 (1997).
- Tans, S. J., Verschueren, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
- Martel, R., Schmidt, T., Shea, H. R., Hertel, T. & Avouris, P. Single- and multi-wall carbon nanotube field-effect transistors. *Appl. Phys. Lett.* **73**, 2447–2449 (1998).
- Joachim, C., Gimzewski, J. K. & Aviram, A. Electronics using hybrid-molecular and mono-molecular devices. *Nature* **408**, 541–548 (2000).
- Andres, R. P. et al. "Coulomb staircase" at room temperature in a self-assembled molecular nanostructure. *Science* **272**, 1323–1325 (1996).
- Dunbar, T. D. et al. Combined scanning tunneling microscopy and infrared spectroscopic characterization of mixed surface assemblies of linear conjugated guest molecules in host alkanethiolate monolayers on gold. *J. Phys. Chem. B* **104**, 4880–4893 (2000).
- Chen, J. et al. Room-temperature negative differential resistance in nanoscale molecular junctions. *Appl. Phys. Lett.* **77**, 1224–1226 (2000).
- Metzger, R. M. Electrical rectification by a molecule: The advent of unimolecular electronic devices. *Acc. Chem. Res.* **32**, 950–957 (1999).
- Collier, C. P. et al. A [2]catenane-based solid state electronically reconfigurable switch. *Science* **289**, 1172–1175 (2000).
- Collier, C. P. et al. Electronically configurable molecular-based logic gates. *Science* **285**, 391–394 (1999).
- Ellenbogen, J. C. & Love, J. C. Architectures for molecular electronic computers: 1. Logic structures and an adder designed from molecular electronic diodes. *Proc. IEEE* **88**, 386–426 (2000).
- Horowitz, G. Organic field-effect transistors. *Adv. Mater.* **10**, 365–377 (1998).
- Forrest, S., Burrows, P. & Thompson, M. The dawn of organic electronics. *IEEE Spect.* **37**, 29–34 (2000).
- Di Ventra, M., Pantelides, S. T. & Lang, N. D. The benzene molecule as molecular resonant-tunneling transistor. *Appl. Phys. Lett.* **76**, 3448–3450 (2000).
- Tour, J. M. Molecular electronics. Synthesis and testing of components. *Acc. Chem. Res.* **33**, 791–804 (2000).
- Joachim, C. & Gimzewski, J. K. An electromechanical amplifier using a single molecule. *Chem. Phys. Lett.* **265**, 353–357 (1997).
- Sze, S. M. *Semiconductor Devices: Physics and Technology* 431–486 (Wiley, New York, 1985).
- Rodgers, T. J. & Meindl, J. D. VMOS: High-speed TTL compatible MOS logic. *J. Solid State Circ.* **9**, 239–250 (1974).
- Hergenrother, J. M. et al. in *International Electron Devices Meeting 2000. Technical Digest* 65–68 (IEEE, Piscataway, 2000).
- Hong, S. et al. Molecular conductance spectroscopy of conjugated, phenyl-based molecules on Au(111): the effect of end groups on molecular conduction. *Superlattices Microstruct.* **28**, 289–303 (2000).
- Tour, J. M. Conjugated macromolecules of precise length and constitution. Organic synthesis for the construction of nanoarchitectures. *Chem. Rev.* **96**, 537–553 (1996).
- Dodabalapur, A., Torsi, L. & Katz, H. E. Organic transistors: two-dimensional transport and improved electrical characteristics. *Science* **268**, 270–271 (1995).
- Di Ventra, M., Pantelides, S. T. & Lang, N. D. First-principle calculation of transport properties of a molecular device. *Phys. Rev. Lett.* **84**, 979–982 (2000).
- Natori, K. Ballistic metal-oxide-semiconductor field effect transistor. *J. Appl. Phys.* **76**, 4879–4890 (1994).
- Emberly, E. G. & Kirczenow, G. Multiterminal molecular wire systems: A self consistent theory and computer simulations of charging and transport. *Phys. Rev. B* **62**, 10451–10458 (2000).
- Chen, J. et al. Electronic transport through metal-1,4-phenylene diisocyanide-metal junctions. *Chem. Phys. Lett.* **313**, 741–748 (1999).
- Karl, N. in *Semiconductors* Vol. 17i (eds Madelung, O., Schulz, M. & Weiss, H.) 106–218 (Landolt-Börnstein, New Series, Springer, Berlin, 1985).
- Schön, J. H., Kloc, Ch. & Batlogg, B. Fractional quantum Hall effect in organic molecular semiconductors. *Science* **288**, 2338–2340 (2000).
- Burshtein, Z. & Williams, D. F. Temperature dependence of carrier generation, and transport in para-terphenyl above and below the 180 K phase transition. *J. Chem. Phys.* **68**, 983–988 (1978).

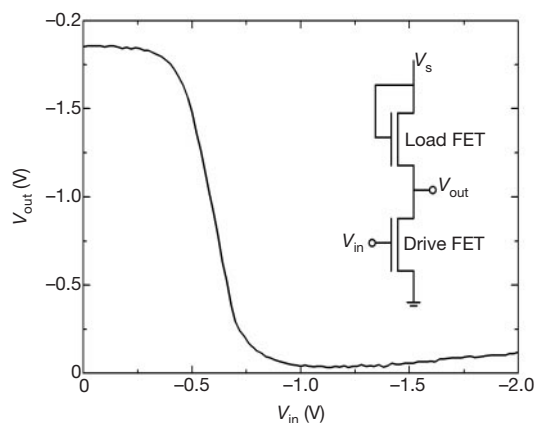


Figure 4 Output characteristics of an inverter using two 4,4'-biphenyldithiol (molecule **2**) SAMFETs. The structure of the inverter circuit is depicted systematically. A gain of 6 is observed for such circuits.

Acknowledgements

We thank D. Abusch-Magder, E. Bucher, F. Capasso, Ch. Kloc, J. A. Rogers, M. Schön and E. A. Chandross for many valuable discussions. J.H.S. thanks E. Bucher for the use of the laboratory during his stay at the University of Konstanz. H.M. is grateful to F. Wudl of the University of California at Los Angeles for his support of an internship at Lucent Technologies.

Correspondence and requests for materials should be addressed to J.H.S. (e-mail: hendrik@lucent.com).

Magnetic carbon

Tatiana L. Makarova^{*,†,‡}, **Bertil Sundqvist**[†], **Roland Höhne**[§],
Pablo Esquinazi[§], **Yakov Kopelevich**^{||}, **Peter Scharff**[‡],
Valerii A. Davydov[¶], **Ludmila S. Kashevarova**[¶]
& **Aleksandra V. Rakhmanina**[¶]

^{*} Ioffe Physico-Technical Institute, 194021 St Petersburg, Russia

[†] Department of Experimental Physics, Umeå University, S-90187 Umeå, Sweden

[‡] Institute for Physics, Chemical Department, TU Ilmenau, D-98693 Ilmenau, Germany

[§] Department for Superconductivity and Magnetism, Leipzig University, D-04103 Leipzig, Germany

^{||} Instituto de Física, Unicamp, 13083-970 Campinas, Sao Paulo, Brazil

[¶] Institute of High Pressure Physics, 142092 Troitsk, Russia

The discovery of nanostructured forms of molecular carbon has led to renewed interest in the varied properties of this element. Both graphite and C_{60} can be electron-doped by alkali metals¹ to become superconducting; transition temperatures of up to 52 K have been attained by field-induced hole-doping of C_{60} (ref. 2). Recent experiments^{3,4} and theoretical studies^{5,6} have suggested that electronic instabilities in pure graphite may give rise to superconducting and ferromagnetic properties, even at room temperature. Here we report the serendipitous discovery of strong magnetic signals in rhombohedral C_{60} . Our intention was to search for superconductivity in polymerized C_{60} ; however, it appears that our high-pressure, high-temperature polymerization process results in a magnetically ordered state. The material exhibits features typical of ferromagnets: saturation magnetization, large hysteresis and attachment to a magnet at room temperature. The temperature dependences of the saturation and remanent magnetization indicate a Curie temperature near 500 K.

Ferromagnetism has previously been observed in two C_{60} compounds: tetrakis(dimethylamino)ethylene-fullerene[60] (TDAE- C_{60} ; refs 7, 8) below 17 K, and 3-aminophenyl-methano-fullerene[60]-cobaltocene below 19 K (ref. 9). It has also been reported that hydrogenated fullerene¹⁰, $C_{60}H_{36}$, and some palladium fullerenes¹¹, $C_{60}Pd_n$, may be ferromagnetic. A ferromagnetic phase of mixed sp^2 and sp^3 pure carbon has been predicted theoretically¹². However, ferromagnetism in organic materials is usually observed only at very low temperatures¹³. The first organic ferromagnet consisting only of light elements¹⁴, $C_{13}H_{16}N_3O_4$, shows long-range ferromagnetic order below 0.65 K, but there are a very few examples of metallo-organic materials with a magnetic ordering temperature exceeding room temperature: V(tetracyanoethylene) $_x$ ·γ(solvent) is a bulk ferrimagnet¹⁵ with a Curie temperature, T_C , estimated as 400 K.

Pristine C_{60} is a van der Waals crystal, which can be converted to covalently bonded crystalline phases by compression¹⁶. Depending on the treatment, the molecules interconnect to form one-, two- or three-dimensional polymers^{17,18}. Here we focus on the two-dimensionally polymerized highly oriented rhombohedral C_{60} phase (Rh- C_{60}), which resembles highly oriented pyrolytic graphite, but with

layers of covalently bonded C_{60} molecules substituted for the graphene layers. If prepared at temperatures below 923 K it is a semiconductor, but an increase in polymerization temperature results in a transition to metallic behaviour¹⁹. The layered structure of the material reveals itself in the anisotropy of electrical properties²⁰.

We produced one-dimensional (orthorhombic) polymers and two-dimensional (rhombohedral and tetragonal) polymers, as well as the phases obtained above the temperature stability limit of the C_{60} cage (about 1,073–1,173 K; ref. 18), and studied their magnetic behaviour. Among them, only the Rh- C_{60} phase shows the ferromagnetic behaviour discussed here. This behaviour is repeatable and stable, and the measurements reported here were made on samples from three different batches, prepared in 1996, 1998 and 2000. Each batch contains samples prepared at different pressures—in particular a series of nine samples prepared at 6 GPa and at temperatures in the range (970–1,170) K $\pm$ 10 K, in steps of 25 K. From these three batches, five of the six samples prepared at 1,025 K or 1,050 K—close to, but below, the stability limit of the fullerene cage—were found to be ferromagnetic with qualitatively similar behaviour.

We characterized the sample structure by Raman spectroscopy, X-ray diffraction²¹ and scanning electron microscopy²². The cell parameters calculated from the X-ray diffraction patterns are $a = 9.204$ Å and $b = 2.461$ Å (space group $R\bar{3}m$). Raman spectra (Fig. 1) indicate an almost single-phase Rh- C_{60} material, with the pentagonal pinch $A_g(2)$ mode, which is quite sensitive to the type and degree of polymerization, peaking at 1,407 cm^{-1} . The mode frequencies obtained agree with those calculated and normally observed for the Rh- C_{60} phase²³, with negligible contributions from other polymeric C_{60} phases and with no indications that the C_{60} cages are damaged.

We observe a difference in the a.c. response for applied magnetic fields H_{ac} parallel to, $\chi_{||}$, and perpendicular to, $\chi_{\perp}$, the polymerized planes. The anisotropy can be understood as being due to the layered structure of the material. The measurements were performed at relatively low a.c. field, and the value of the (dissipative) χ'' was found to be at the limit of the instrument sensitivity. Therefore, we restrict ourselves to showing in Fig. 2 the following features of the (inductive) χ' of Rh- C_{60} : (1) unlike the values found for nearly all other forms of carbon, it is positive; (2) it increases slightly with temperature; (3) it shows an upturn below 30 K; and (4) the ratio $\chi_{\perp}/\chi_{||}$ is 5.5.

Pristine, that is, unpolymerized, C_{60} is diamagnetic with a susceptibility of $\sim 3.4 \times 10^{-7}$ e.m.u. g^{-1} (ref. 24). The paramagnetic

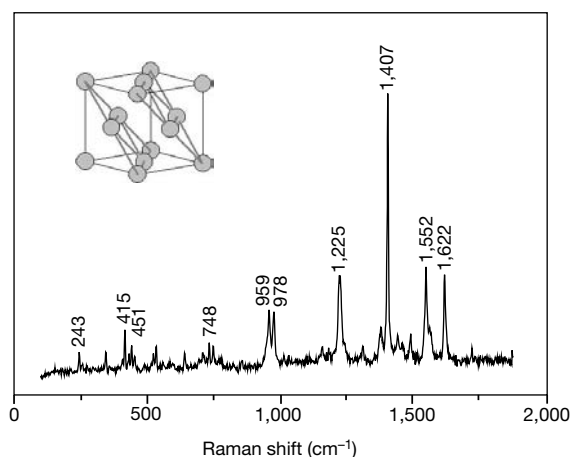


Figure 1 Raman spectrum of magnetically ordered rhombohedral C_{60} polymer (Rh- C_{60}). Inset, illustration of the relation between the structures of the pressure-induced polymer and the original cubic lattice of the pristine C_{60} .

upturn at low temperatures is always observed in the a.c. susceptibility of pristine C_{60} (ref. 25), except for specially prepared oxygen-free C_{60} single crystal. Thus the upturn that we observe below 30 K can be attributed to a superimposed Curie term due to oxygen that entered the crystals during loading of the samples into the pressure cell.

Magnetic anisotropy has been observed in structurally anisotropic forms of carbon, namely nanotubes and graphite. The magnetic response along the nanotubes is 10% less diamagnetic than that perpendicular to the tubes²⁶. The magnetic susceptibility of graphite is large and diamagnetic, $\chi_{\parallel} < -5 \times 10^{-7}$ e.m.u. g⁻¹, $\chi_{\perp} = -5 \times 10^{-5}$ e.m.u. g⁻¹. A transition from diamagnetic to paramagnetic behaviour has been observed in acceptor intercalation compounds and fluorinated graphite. This paramagnetism is caused by local magnetic moments of dangling bonds²⁷. A paramagnetic response with a large Curie contribution has also been observed for the one-dimensional polymeric phase of RbC_{60} (ref. 28). Undoped $Rh-C_{60}$ also shows paramagnetic behaviour, and we emphasize that the magnitudes of the susceptibility are 100 times larger than for graphite and 10,000 times larger than for pristine C_{60} .

We performed SQUID (superconducting quantum interference device) magnetometry on $Rh-C_{60}$, and found a magnetically ordered phase. Figure 3a shows magnetization hysteresis loops measured in the field range $-2 \text{ kOe} < H < 2 \text{ kOe}$, for temperatures of 10 K and 300 K. Figure 3b shows these two loops over a wider range of fields, where saturation of the magnetization is clearly seen above $\sim 2 \times 10^4$ Oe. A relatively small diamagnetic contribution ($M = \chi H$, with χ values of -1.2×10^{-6} e.m.u. g⁻¹ Oe⁻¹ at 10 K, and -1.4×10^{-6} e.m.u. g⁻¹ Oe⁻¹ at 300 K) was subtracted from the original data. Using the spin concentration value obtained from electron spin resonance data $n = 5 \times 10^{18} \text{ cm}^{-3}$, we estimate the magnetic moment per electron as $0.4\mu_B$, where μ_B is the Bohr magneton.

In Fig. 4 we show the temperature dependence of the magnetization at a fixed field of 0.2 T, and also the temperature dependence of the remanent magnetization obtained at $H = 0$ after decreasing the applied field from 2,000 Oe. A clearly defined ferromagnetic transition is observed with a Curie temperature of $T_C \approx 500$ K. The transition is reversible. The loops were preserved after annealing at 520 and 640 K for 2 hours, which indicates phase stability.

Such well defined ferromagnetic behaviour has not, to our knowledge, been previously observed in pure carbon. Some of our samples show magnetization that is strong enough for them to be lifted off a table surface by a small SmCo magnet (see Supplementary Information). After our initial discovery, we performed a number of additional experiments on a specially synthesized batch of samples. We took great care to eliminate all traces of

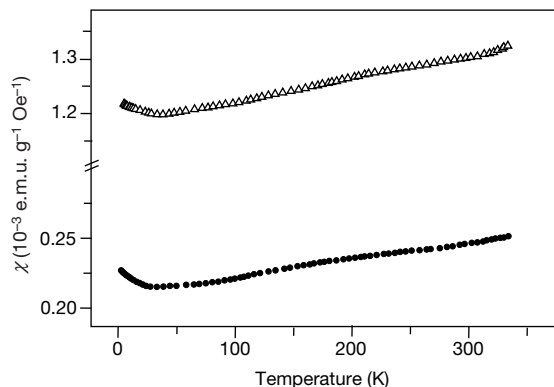


Figure 2 Temperature dependence of the magnetic susceptibility of $Rh-C_{60}$. Upper curve, triangles: magnetic field normal to the polymerized planes. Lower curve, circles: magnetic field directed along the planes.

impurities, and magnetic measurements were also carried out on the pristine unpolymerized material. After measurements, some pieces of ferromagnetic polymeric samples were depolymerized by heating to 700 K for several hours, which resulted in complete loss of ferromagnetism. The measurements on this new batch of samples clearly showed that the results were repeatable, but also indicated that not all nominally rhombohedral samples showed ferromagnetic behaviour: the effect is very sensitive to the pressure and temperature used in sample preparation.

The purity of the material is critical, as there have been several false reports on 'organic magnets' that included trace amounts of ferromagnetic metal. The most convincing evidence we have against impurity effects is that neither the pristine fullerene powder nor depolymerized samples show any ferromagnetic behaviour. We have paid great attention to chemical analysis of the pristine material as well as of the polymerized phase. The total amount of magnetic (Fe, Ni, Co) impurities is 22 p.p.m. in the pristine phase. The saturation magnetic moment per unit mass caused by these impurities is estimated as $0.003 \text{ e.m.u. g}^{-1}$, that is, 30 times less than the magnetization measured in our experiments. This excludes the possibility that the effect is due to ferromagnetic clusters formed by magnetic atoms. A bulk impurity spin ordering is very unlikely, owing to the large average distance between the magnetic impurities. We conclude that the clear observations of ferromagnetism shown in Figs 3 and 4 can be explained only by intrinsic properties of the C_{60} polymer, such as ordered defect moments or itinerant ferromagnetism.

As long as the magnetic properties arise from the C_{60} polymer itself, two facts need to be taken into account. First, unpaired spins can result from polymerization but, second, they result from deviations from the conventional polymerization process. Available models for the polymerization of neutral molecules do not predict the appearance of unpaired electrons. The covalent bonds are formed between two hexagon facets via the 2+2-cycloaddition mechanism²⁹, leading to a sp^2-sp^3 rehybridization. Diamond-like bonds could be said to form, but this is not strictly accurate because

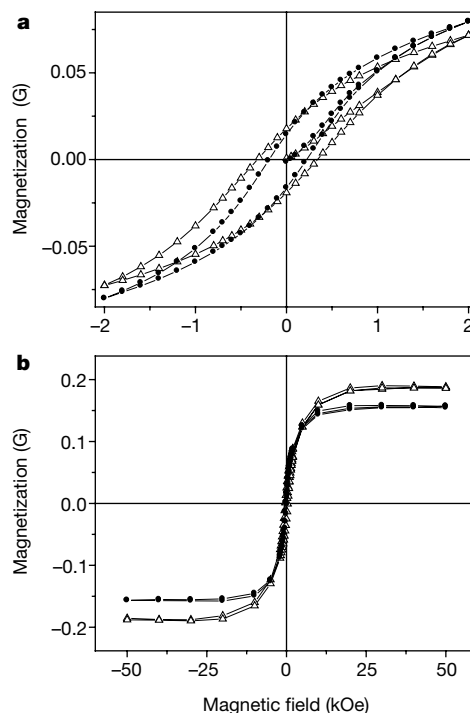


Figure 3 Hysteresis loops for $Rh-C_{60}$. Data were obtained at $T = 10$ K (triangles) and 300 K (circles). **a**, Hysteresis is observed in the field range $-2 \text{ kOe} < H < 2 \text{ kOe}$; **b**, saturation of the magnetization is clearly seen in a broader field region.

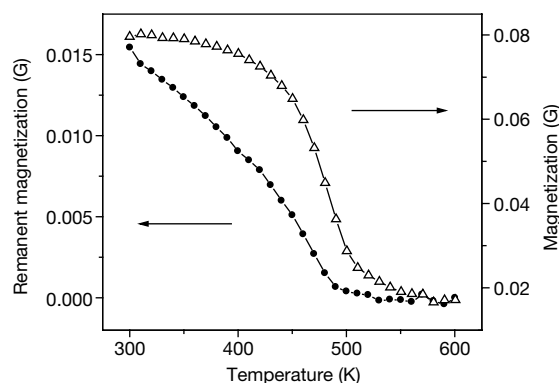


Figure 4 Magnetization of Rh-C₆₀ in a fixed applied field of 0.2 T (upper curve, triangles) and the remanent magnetization obtained at $H = 0$ T (lower curve, circles) as a function of temperature. The Curie temperature is about 500 K.

the bond angle in diamond is 109.42°, whereas the interfullerene bonds are directed at right angles. The bond dissociation energy is 1.9 eV, which is only one-half of the bond energy for diamond. We suggest that the σ -bond is formed by one electron, and another electron is transferred to a higher orbital and can participate in the magnetic ordering.

Another possibility is the presence of defect patterns in the polymerized network. Fullerene C₆₀ has two symmetry-independent bonds: one separates two hexagons and another lies between a hexagon and a pentagon. When a misoriented C₆₀ molecule is inserted into the layer, a hexagon/pentagon bond can be involved in polymerization, producing a gapless density of states with a conduction electron³⁰. Itinerant ferromagnetism can therefore originate from topological defects in the polymerized layer. When considering possible models for magnetism in Rh-C₆₀, we must also take into account the fact that strong magnetic properties are only observed in structures prepared at 1,020–1,050 K—that is, close to the molecule stability limit. X-ray diffractograms and Raman spectra have shown no signs of amorphous carbon, but we cannot exclude the possibility that the magnetic effects are caused by the deformation and partial destruction of the fullerene cage.

We seem to be observing the interplay of two processes. On the one hand, an increase in the preparation temperature results in an increased degree of polymerization, which is characterized by a certain mutual orientation of fullerene molecules. The relative orientation of the C₆₀ molecules has been proved to govern the magnetic state of TDAE-C₆₀: it can exist in either a ferromagnetic, a superparamagnetic or a spin-glass state⁷. On the other hand, the magnetic effects occur in the vicinity of the temperature at which the cage collapses with the formation of an amorphous carbon phase, and these samples may contain many defects, which would cause self-doping. We are at present performing a detailed comparative study of the structural, magnetic and transport properties of Rh-C₆₀ synthesized under different conditions, in an attempt to determine more precisely the causes of the magnetic behaviour reported here. □

Methods

Undoped polymerized C₆₀ phases were produced in a toroid-type high-pressure apparatus. The preparation temperature varied from 273 to 1,123 K at pressures of 1.5–6 GPa. The starting C₆₀ material, supplied by Term USA, was the same for all the samples. High-purity (99.999%), twice-sublimed, C₆₀ powder of small crystal size, with a total amount of metallic impurities of ~22 p.p.m., was stored in vacuum-sealed ampoules, which were opened just before the high-pressure treatment. Cylindrical samples, prepared by cold pressing of the initial powder, were wrapped in Nb foil and placed in a boron nitride cage, which screened the samples from the graphite heater. The samples were compressed, heated and held isothermally, then quickly cooled (quenched) to conserve the phases obtained.

Impurity elements were determined by graphite furnace atomic absorption spectrometry using the slurry technique; a Perkin-Elmer SIMAA 6000 with Zeeman-effect background correction was used. The complex a.c. magnetic susceptibility was measured in a field of 5 Oe and in zero bias field between 2 and 350 K at a frequency of 1,000 Hz using the mutual-inductance method with a lock-in amplifier (Oxford MagLab 2000 System). The a.c. field was applied either perpendicular or parallel to the crystalline planes. The d.c. magnetization measurements were performed with a SQUID magnetometer in the range 4–640 K and in fields up to 50 kOe. The sample (mass 3.2 mg) was fixed between two quartz props in the centre of a quartz tube with a closed end. The quartz tube was vacuum sealed at room temperature and at a pressure of 2×10^{-6} mbar.

Received 9 May; accepted 30 August 2000.

- Hebard, A. F. *et al.* Superconductivity at 18 K in potassium-doped C₆₀. *Nature* **350**, 600–601 (1991).
- Schön, J. H., Kloc, C. & Batlogg, B. Superconductivity at 52 K in hole-doped C₆₀. *Nature* **408**, 549–552 (2000).
- Kopelevich, Y., Esquinazi, P., Torres, J. H. S. & Moehlecke, S. Ferromagnetic- and superconducting-like behavior of graphite. *J. Low Temp. Phys.* **119**, 691–702 (2000).
- Kempa, H. *et al.* Magnetic-field-driven superconductor-insulator-type transition in graphite. *Solid State Commun.* **115**, 539–542 (2000).
- González, J., Guinea, F. & Vozmediano, M. A. H. Electron-electron interactions in graphene sheets. *Phys. Rev. B* **63**, 134421–1–134421-8 (2001).
- Hirigaya, K. The mechanism of magnetism in stacked nanographite: theoretical study. *J. Phys. Condens. Matter* **13**, 1295–1302 (2001).
- Narymbetov, B. *et al.* Origin of ferromagnetic exchange interactions in a fullerene-organic compound. *Nature* **407**, 883–885 (2000).
- Allemand, P. M. *et al.* Organic molecular soft ferromagnetism in a fullerene C₆₀. *Science* **253**, 301–303 (1991).
- Mrzel, A. *et al.* Ferromagnetism in a cobaltocene-doped fullerene derivative below 19 K due to unpaired spins only on fullerene molecules. *Chem. Phys. Lett.* **298**, 329–334 (1998).
- Lobach, A. S. *et al.* C₆₀H₁₈, C₆₀H₃₆ and C₇₀H₃₆ fullerene hydrides: Study by methods of IR, NMR, XPS, EELS and magnetochemistry. *Fullerene Sci. Technol.* **6**, 375–391 (1998).
- Shul'ga, Y. M. *et al.* Magnetic properties of C₆₀Pd₉ fullerides. *Mol. Cryst. Liq. Cryst.* **10**, 201–206 (1998).
- Ovchinnikov, A. A. & Shamovsky, I. L. The structure of the ferromagnetic phase of carbon. *J. Mol. Struct. (Theorchem)* **83**, 133–140 (1991).
- Miller, J. S. & Epstein, A. J. Molecule-based magnets—An overview. *Mater. Res. Soc. Bull.* **25**, 21–28 (2000).
- Kahashi, M. *et al.* Discovery of a quasi-1D organic ferromagnet, P-NPN. *Phys. Rev. Lett.* **67**, 746–748 (1991).
- Du, G., Joo, J., Epstein, A. J. & Miller, J. S. Anomalous charge transport phenomena in molecular-based magnet V(TCNE)_x·y(solvent). *J. Appl. Phys.* **73**, 6566–6568 (1993).
- Iwasa, Y. *et al.* New phases of C₆₀ synthesised at high pressure. *Science* **264**, 1570–1572 (1994).
- Núñez-Requero, M., Marques, L., Hodeau, J. L. & Perroux, M. Polymerized fullerite structures. *Phys. Rev. Lett.* **74**, 278–281 (1995).
- Sundqvist, B. Fullerenes under high pressures. *Adv. Phys.* **48**, 1–134 (1999).
- Makarova, T. L. *et al.* Anisotropic metallic properties of highly-oriented C₆₀ polymer. *Synth. Met.* **121**, 1099–1100 (2001).
- Makarova, T. L. *et al.* Electrical properties of two-dimensional fullerene matrices. *Carbon* **39**, 2203–2209 (2001).
- Makarova, T. L. *et al.* in *Fullerenes Vol. 7, Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M., Guldi, D. M. & Kamat, P. V.) 628–639 (Proceedings series PV 99-12, Electrochemical Society, Pennington, New Jersey, 1999).
- Makarova, T. L. *et al.* Conductivity of two-dimensional C₆₀ polymers. *Mol. Cryst. Liq. Cryst. C* **13**, 151–156 (2000).
- Davydov, V. A. *et al.* Spectroscopic study of pressure-polymerized phases of C₆₀. *Phys. Rev. B* **61**, 11936–11945 (2000).
- Luo, W. L., Wang, H., Ruoff, R. C., Cioslowski, J. & Phelps, S. Susceptibility discontinuity in single-crystal C₆₀. *Phys. Rev. Lett.* **73**, 186–188 (1994).
- Haddon, R. C. *et al.* Experimental and theoretical determination of the magnetic susceptibility of C₆₀ and C₇₀. *Nature* **350**, 46–47 (1991).
- Wang, X. K., Chang, R. P. H., Patashinski, A. & Ketterson, J. B. Magnetic susceptibility of buckytubes. *J. Mater. Res.* **9**, 1578–1582 (1994).
- Panich, M., Shames, A. I. & Nakajima, T. On paramagnetism in fluorinated graphite: EPR and solid state NMR study. *J. Phys. Chem. Solids* **62**, 959–964 (2001).
- Saito, T., Akita, Y. & Tanaka, K. Magnetic susceptibility of the one-dimensional polymeric phase of RbC₆₀. *Phys. Rev. B* **61**, 16091–16096 (2000).
- Pekker, S., Forro, L., Mihaly, L. & Janossy, A. Orthorhombic A₁C₆₀—a conducting linear alkali fulleride polymer. *Solid State Commun.* **90**, 349–352 (1994).
- Xu, C. H. & Scuseria, G. E. Theoretical predictions for a 2-dimensional rhombohedral phase of solid C₆₀. *Phys. Rev. Lett.* **74**, 274–277 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements

We thank R. Blinc for discussions, W. Frech for help with impurity analysis, and A. L. Shelankov for comments on the manuscript. This work was done in the framework of a project supported by the Alexander von Humboldt Foundation (T.L.M.). Part of this work was supported by the DFG and FAPESP. B.S. thanks the Swedish Research Council for support.

Correspondence and requests for materials should be addressed to T.L.M. (e-mail: tatiana.makarova@physics.umu.se).

Orbitally induced oscillations in the East Antarctic ice sheet at the Oligocene/Miocene boundary

Tim R. Naish¹, Ken J. Woolfe^{2,3}, Peter J. Barrett⁴, Gary S. Wilson⁵, Cliff Atkins⁴, Steven M. Bohaty⁶, Christian J. B  cker^{7,8}, Michele Claps^{9,8}, Fred J. Davey¹, Gavin B. Dunbar^{2,4}, Alistair G. Dunn¹⁰, Chris R. Fielding¹¹, Fabio Florindo^{12,13}, Michael J. Hannah⁴, David M. Harwood⁶, Stuart A. Henrys¹, Lawrence A. Krissek¹⁴, Mark Lavelle¹⁵, Jaap van der Meer^{16,8}, William C. McIntosh¹⁷, Frank Niessen¹⁸, Sandra Passchier¹⁴, Ross D. Powell¹⁹, Andrew P. Roberts¹³, Leonardo Sagnotti¹², Reed P. Scherer¹⁹, C. Percy Strong¹, Franco Talarico²⁰, Kenneth L. Verosub²¹, Giuliana Villa²², David K. Watkins⁶, Peter-N. Webb¹⁴ & Thomas Wonik⁷

Between 34 and 15 million years (Myr) ago, when planetary temperatures were 3–4°C warmer than at present and atmospheric CO₂ concentrations were twice as high as today¹, the Antarctic ice sheets may have been unstable^{2–7}. Oxygen isotope records from deep-sea sediment cores suggest that during this time fluctuations in global temperatures and high-latitude continental ice volumes were influenced by orbital cycles^{8–10}. But it has hitherto not been possible to calibrate the inferred changes in ice volume with direct evidence for oscillations of the Antarctic ice sheets¹¹. Here we present sediment data from shallow marine cores in the western Ross Sea that exhibit well dated cyclic variations, and which link the extent of the East Antarctic ice sheet directly to orbital cycles during the Oligocene/Miocene transition (24.1–23.7 Myr ago). Three rapidly deposited glaci-marine sequences are constrained to a period of less than 450 kyr by our age model, suggesting that orbital influences at the frequencies of obliquity (40 kyr) and eccentricity (125 kyr) controlled the oscillations of the ice margin at that time. An erosional hiatus covering 250 kyr provides direct evidence for a major episode of global cooling and ice-sheet expansion about 23.7 Myr ago, which had previously been inferred from oxygen isotope data (Mil event⁵).

Sediment cores recovered from the Antarctic continental margin during the past three decades have indicated that grounded ice has covered much of Antarctica since Oligocene times^{2–4,12}, with the earliest ice sheets forming close to the Eocene/Oligocene boundary around 34 Myr ago^{4–7}. Oligocene strata in the longest and most complete of these cores (CIROS-1) recorded cycles of ice-margin

advance and retreat representing ice-volume changes that might have driven past sea-level fluctuations³, such as those later inferred from high-resolution, deep-sea oxygen isotope records^{5,9,10}. However, the frequency of these oscillations could not be determined. This was because core stratigraphy and sedimentology indicated numerous potential time breaks, and because core chronologies, based largely on biostratigraphic methods, allowed a time resolution of no better than a million years.

In the austral summers of 1998 and 1999, the Cape Roberts Project (CRP) drilled 1,500 m of strata from the western margin of the Victoria Land basin^{13–15} (Fig. 1); 46 unconformity-bound, Oligocene to Early Miocene (33–17 Myr ago) glaci-marine cycles, or depositional sequences, were recorded. The strata accumulated some kilometres off the coast as part of a laterally extensive seaward-thickening nearshore wedge. This geometry is evident in seismic profiles oriented parallel to and normal to the coast¹⁶. During periods of glacial advance, the interior ice sheet fed through outlet glaciers to a laterally extensive marine ice terminus that extended well out onto the continental shelf beyond the CRP drill sites. In periods of ice retreat the drill sites lay off an open wave-dominated coast, with deposition of mud, and occasional debris from floating ice. Thus cycles of ice-margin advance and retreat across the shelf are viewed as cycles of expansion and contraction of the East Antarctic ice sheet recorded as unconformity-bound sequences. This view is supported by the style of erosion within these strata—laterally extensive, sub-horizontal erosion surfaces, judging from coast-parallel seismic sections.

Each sequence (Fig. 2) begins with an erosion surface followed by a characteristic vertical facies succession, interpreted to represent: (1) erosion during ice-margin advance followed by deposition during ice-margin retreat, (2) relatively ice-free open marine sedimentation and (3) re-advance of the ice margin into a shallow marine setting. Sedimentary features of this succession are given in more detail below, but it is important to note that sedimentological and faunal palaeobathymetric indicators within cycles suggest that water depth changes of up to ~50 m occurred in concert with ice-margin advance and retreat cycles^{13,14}. Variations in grain size and changes in lithofacies with time primarily reflect oscillations in depositional energy that were most probably controlled by the combination of changes in shoreline and glacial proximity. Such

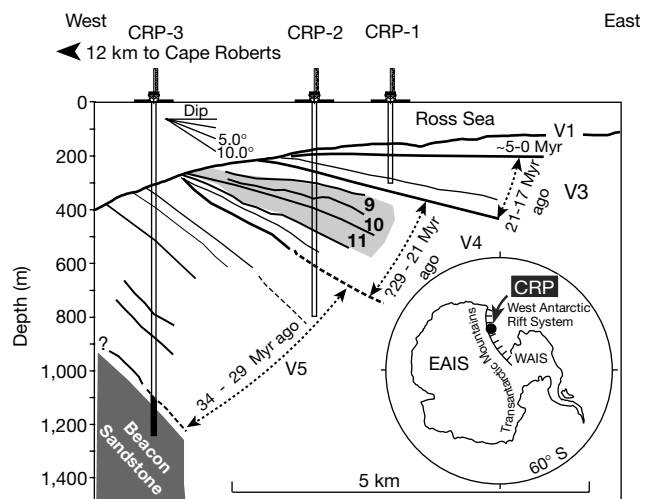


Figure 1 Cross-section through Cenozoic strata beneath the western flank of Roberts ridge, western Ross Sea. The geometry and age of strata recovered by CRP drilling is shown. Depositional sequences 9, 10 and 11 in CRP-2/2A, the focus of this study, are mappable in seismic reflection profiles¹⁶. EAIS, East Antarctic ice sheet; WAIS, West Antarctic ice sheet. V1–5 refer to Ross Sea seismic units (see ref. 16).

¹ Institute of Geological and Nuclear Sciences, PO Box 30368, Lower Hutt, New Zealand; ² School of Earth Sciences, James Cook University, Townsville, Queensland 4811, Australia; ³ Antarctic Research Centre, Victoria University of Wellington, PO Box 600, Wellington, New Zealand; ⁴ Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK; ⁵ Department of Geosciences, University of Nebraska, Lincoln, Nebraska 68588-0340, USA; ⁶ Institut für Geowissenschaftliche Gemeinschaftsaufgaben, Stilleweg 2, D-30655 Hannover, Germany; ⁷ Istituto di Scienze del Mare, Università degli Studi di Ancona, Via Brece Bianche, 60131 Ancona, Italy; ⁸ National Institute of Water and Atmospheric Research, PO Box 14901, Wellington, New Zealand; ⁹ Department of Earth Sciences, University of Queensland, Brisbane, Queensland 4072, Australia; ¹⁰ Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143 Roma, Italy; ¹¹ School of Ocean and Earth Science, University of Southampton, Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK; ¹² Byrd Polar Research Centre and Department of Geological Sciences, The Ohio State University, Columbus, Ohio 43210, USA; ¹³ British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK; ¹⁴ Fysisch Geografisch en Bodemkundig Laboratorium, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands; ¹⁵ New Mexico Institute of Mining and Technology, Socorro, New Mexico 87801, USA; ¹⁶ Alfred-Wegener-Institute, Postfach 12 01 81, 27516 Bremerhaven, Germany; ¹⁷ Department of Geology and Environmental Geosciences, Northern Illinois University, De Kalb, Illinois 60115, USA; ¹⁸ Dipartimento di Scienze della Terra, Università degli Studi di Siena, Via del Laterano 8, 53100, Siena, Italy; ¹⁹ Department of Geology, University of California, Davis, California 95616, USA; ²⁰ Dipartimento di Scienza della Terra, Università degli Studi di Parma, Parco Area Scienze 157A, 43100 Parma, Italy

²¹ Deceased.

²² Present addresses: RWE-DEA AG, Ueberseering 40 22297 Hamburg, Germany (C.J.B.); Agip Kazakhstan North Caspian Operating Company, Raamweg 26, 2596 HL, The Hague, The Netherlands (M.C.); Department of Geography, Queen Mary, University of London, Mile End Road, London E1 4NS, UK (J.v.d.M.).

inferences are consistent with established depositional models of temperate and polythermal (sub-polar) glaciers entering the sea¹⁷, and with models of shallow marine sedimentation during orbitally driven cyclical changes in relative sea level¹⁸.

Basal sequence boundaries are overlain by massive diamictite and sandstone. Clast fabrics are weakly oriented or random, suggesting a large component of rainout debris, which in many cases may have been remobilized in gravity flows. Subglacial shear is indicated by syndepositional soft sediment deformation structures, intraclasts, and clastic dykes within diamictites¹⁹. Subglacial features are indicated by 'turbate' and 'linear' microstructures and a pervasive 'plasmic' fabric, characterized by an oriented clay matrix, within diamictites and deformed sediments beneath diamictites²⁰. The basal diamictite of sequence 11 (for example, Fig. 2) displays synsedimentary shearing and contains striated clasts¹⁴. Sequence boundaries are formed by erosional processes during ice advance and the overlying sediments, with their lack of ice-contact features and fining-upwards texture, are mostly the product of a retreating ice margin.

Basal sediments are followed by a sparsely fossiliferous, bioturbated shelf mudstone corresponding to an interval of maximum water depth, the glacial minimum, and the lowest sedimentation rate within a cycle. A sharp-based, shallow-marine, regressive sandstone facies succession occurs in the upper parts of most sequences. The shallowing character of this succession, and the truncation by the overlying sequence boundary, are both consistent with sediments deposited during a period of relative sea-level fall followed by erosion by the re-advancing ice margin.

Figure 2 shows systematic variations in grain size, clast content, diatom abundance and γ -radiation for one of the more fully developed depositional sequences in CRP-2/2A drill core (Fig. 1), sequence 11. These variations represent environmental proxies that provide additional support to the ice margin and relative water depth interpretations outlined above. The up-section decrease in sand content, clast abundance²¹ (ice-rafted debris), and diatom abundance²² in the lower part of sequence 11 is consistent with

progressive withdrawal of the ice margin from the Ross Sea in concert with a decrease in hydraulic energy during shoreline transgression and relative sea-level rise. The main peak in the number of *in situ* diatom tests corresponds to a mudstone interval within sequence 11, interpreted as ice-distal open shelf and indicative of increased oceanic connection during interglacials. In contrast, peaks in reworked diatom assemblages correspond to diamictites. The γ -ray log displays a positive shift from quartz- and plagioclase-rich, ice-proximal sandstone upward into an interval of high-potassium, illite-dominated, ice-distal shelf mudstone²³.

Most of the time interval spanned by the Cape Roberts drillcore is not represented by the stratigraphic record owing to non-deposition and erosion through long-term tectonic influences and shorter-term glacial processes^{14,15}. However, intervals of relatively continuous deposition occur where increased rates of basin subsidence provided sufficient accommodation space. Thus, although fragmentary, the CRP record provides high-resolution windows into the history and dynamics of the East Antarctic ice sheet. One such interval includes sequences 9, 10 and 11 (Fig. 1) in the CRP-2/2A core. These sequences are on average much thicker than the rest, and include a number of high-precision age datums (polarity reversals and numeric ages on volcanic ashes) providing a chronology that allows estimation of sequence durations at Milankovitch periodicities (Fig. 3). This chronology, which is based on a combination of single-crystal $^{40}\text{Ar}/^{39}\text{Ar}$ ages²⁴, microfossil biostratigraphy^{22,25}, $^{87}\text{Sr}/^{86}\text{Sr}$ analyses²⁶, and correlation of a magnetic polarity zonation to the magnetic polarity timescale (MPTS)²⁷, permits the construction of a robust age model for sequences 9 to 11²⁸.

Single-crystal, laser-fusion $^{40}\text{Ar}/^{39}\text{Ar}$ tephra ages at 280 and 193 m below sea floor (m.b.s.f.), $^{87}\text{Sr}/^{86}\text{Sr}$ ages on *in situ* molluscan samples at 247 and 195 m.b.s.f., and the last occurrence (LO) of *Lisitzinia ornata* at 255 m.b.s.f. (in C6Cr in Southern Ocean cores²⁹) constrain correlation of the normal-polarity interval (magnetozone N5) containing sequences 10 and 11 with geomagnetic chron C6Cn.3n. Stratigraphically higher magnetozone R4 and N4 comprising the lower three-quarters of sequence 9 are correlated with

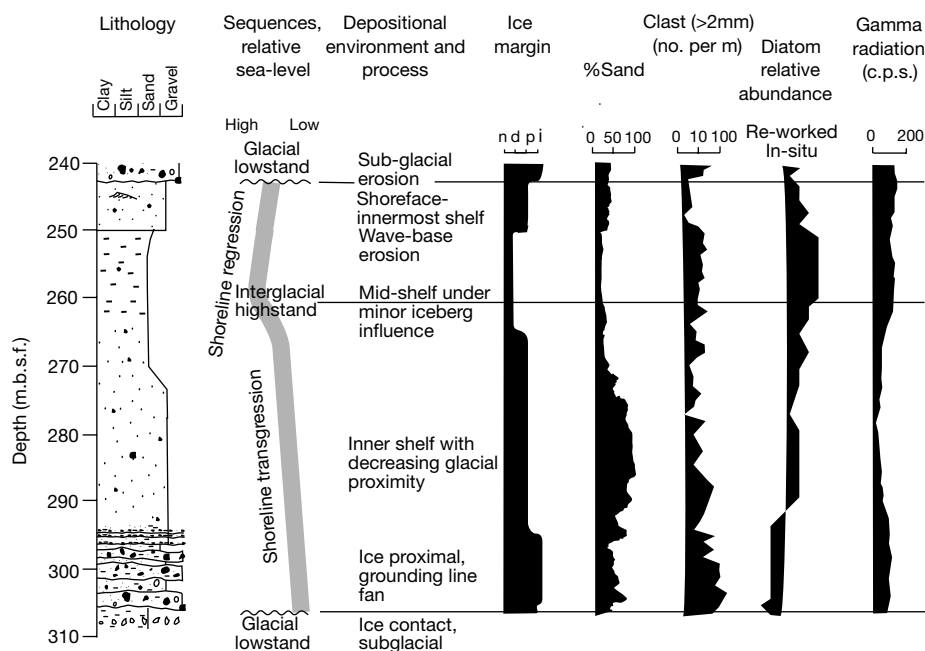


Figure 2 Lithological and environmental changes through sequence 11. Vertical variations in grain size, ice-rafted debris, diatom abundance and γ -radiation are shown for a more fully developed orbitally driven, ice advance–retreat–readvance cycle in

CRP-2/2A. Key for the interpretation of ice margin proximity: n, non-glacial (marine); d, distal glacimarine; p, proximal glacimarine; i, ice contact¹⁴.

C6Cn.2r and C6Cn.2n, respectively. The upper quarter of sequence 9, which has reversed polarity (magnetozones R3) and contains the LO of *Dictyococcites bisectus*, is correlated with earliest C6Cn.1r³⁰. The age model implies an age range from 24.12 to 23.68 Myr. Given that the thin magnetically reversed interval (magnetozones R4) corresponds to rapidly deposited ice-proximal diamictite and sandstone at the base of sequence 9, most of the sedimentary record of C6Cn.2r is considered to have been lost at the underlying sequence boundary—perhaps as much as 250 kyr. The Oligocene/Miocene boundary is placed within sequence 9 at the base of C6Cn.2n at 183.70 m.b.s.f.²⁸.

An alternative age model has been proposed²⁸ that correlates sequence 9 (magnetozones R4, N4 and R3) with the upper part of C6Cn.3n. In this age model, the 4-m-thick reversed polarity interval of diamictite and sandstone at the base of sequence 9 is inferred to represent a short-duration geomagnetic excursion within C6Cn.3n, and the LO of *D. bisectus* in CRP-2/2A is correlated with C6Cn.2r in the earliest Miocene. This alternative age model has been considered unlikely²⁸ because: (1) the polarity signal is robust for the interval of CRP core, (2) the MPTS³¹ contains no evidence of 'tiny wiggles' at

this time and sedimentary records from ODP site 1090 spanning C6Cn.3n show no indication of short polarity intervals (J. Channell, personal communication), and (3) the age of the LO of *D. bisectus* would be inconsistent with Southern Ocean records³⁰.

Our preferred age model restricts the duration of the three glacial sequences to no more than ~450 kyr. Sequence 9 is of ~125 kyr duration (C6Cn.2n spans 123 kyr), and may reflect orbital control of the East Antarctic ice sheet at the eccentricity frequency. The stratigraphically underlying sequences 10 and 11 are restricted to C6Cn.3n, a duration of 119 kyr. Moreover, with sequence 10 appearing to be truncated¹³ and with most of C6Cn.2r inferred to be missing at the sequence 9/10 unconformity, we contend that sequences 10 and 11 probably reflect 40-kyr, obliquity-controlled oscillations in the margin of the East Antarctic ice sheet. On this basis, we present a correlation of sequences 9, 10 and 11 with a high-resolution, Late Oligocene–Early Miocene deep-sea oxygen isotope record from western Atlantic, ODP site 929^{9,10} (Fig. 4).

Figure 4 shows a detailed characterization of the Mi1 oxygen isotope excursion⁵, which is defined by an overall ~1.5‰ increase

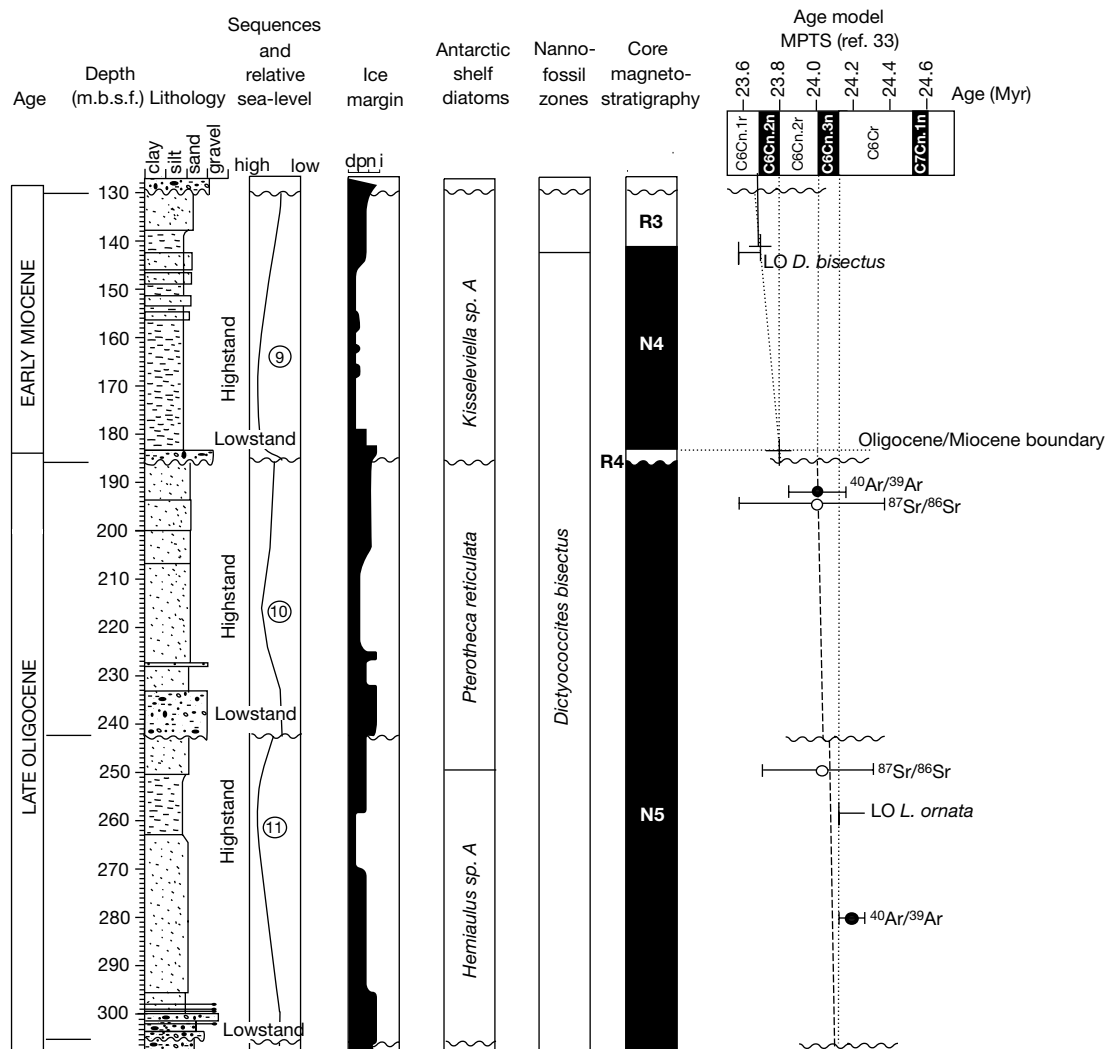


Figure 3 Lithostratigraphy, sequence stratigraphy, variation in sand content and age models for sequences 9, 10 and 11 (130.27–306.65 m) from the CRP-2/2A drill hole. Key for ice margin proximity is shown in Fig. 2 legend. Datums include: (1) the LO (last occurrence) of *Dictyococcites bisectus* at 144.44 m and LO of *Lisitzinia ornata* at 259.2 m in C6Cn.1r and C6Cr, respectively, (2) the ⁴⁰Ar/³⁹Ar pumice dates at 193.40 m

(23.98 ± 0.13 Myr) and at 280.03 m (24.22 ± 0.03 Myr) and (3) the ⁸⁷Sr/⁸⁶Sr dates on *in situ* mollusca at 194.87 m (24.00 ± 0.4/–0.5 Myr) and 247.00 m (24.02 ± 0.35 Myr). The length of vertical bars on biostratigraphic datums indicates the range of error from the respective sampling interval on which the datums are defined.

in $\delta^{18}\text{O}$, equivalent to a sea-level lowering of ~ 80 m accompanied by a decrease in deep-water temperature of 3°C (refs 9, 10). The increase and subsequent decrease in oceanic mean $\delta^{18}\text{O}$ across the Mi1 event has been interpreted as a long-term, 400-kyr, eccentricity cycle⁹. Higher-frequency 100-kyr (eccentricity) and 40-kyr orbital (obliquity) components punctuate the Mi1 excursion. Strong covariance reported in the benthic and planktonic foraminiferal $\delta^{18}\text{O}$ record and the absence of a Northern Hemisphere ice sheet is taken to suggest the presence of retreating and advancing ice sheets on Antarctica⁹. Assuming a Pleistocene calibration of 0.11‰ $\delta^{18}\text{O}$ per 10 m of sea-level, sea-level fluctuations of ~ 30 – 40 m are indicated at 40- and 100-kyr frequencies with corresponding changes in deep-water temperature of 1 – 2°C . Like the authors of ref. 10, we find it intriguing that the climatic deterioration leading into the Mi1 event is punctuated by 40-kyr cycles, yet 100-kyr cycles dominate the $\delta^{18}\text{O}$ record after the Mi1 excursion. Although not well understood, certain periods of Earth history (for example, the past 800 kyr) have been dominated by a climatic response at 100-kyr orbital frequencies. This phenomenon has been explained by several mechanisms involving nonlinear responses to orbital forcing—these include feedbacks in ocean–atmosphere circulation, the global carbon cycle, and the internal dynamics and bedrock interactions associated with large continental ice sheets in polar regions³¹.

In summary, the pattern of ice-volume change that is indicated by the deep ocean record (Fig. 4) is consistent with the pattern of ice marginal sedimentation on Antarctica implied from our analysis of the CRP-2/2A core for the interval 24.1–23.7 Myr ago. Oscillations in ice margin inferred from glacial sequences 11 and 10 can be correlated with 40-kyr, $\delta^{18}\text{O}$ cycles at the onset of the Mi1 excursion.

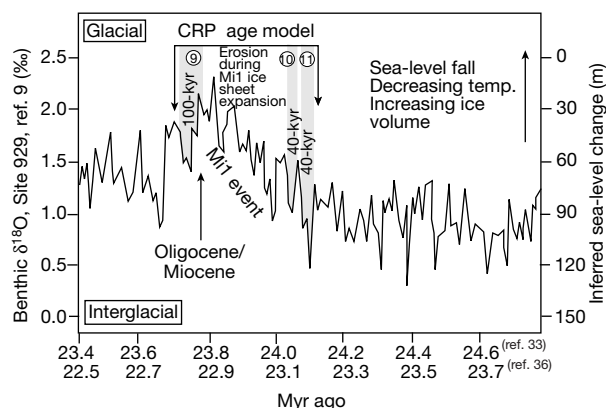


Figure 4 Correlation of sequences 9, 10 and 11 in CRP-2/2A core, western Ross Sea with deep-water benthic foraminiferal $\delta^{18}\text{O}$ record of ODP site 929, western equatorial Atlantic. The $\delta^{18}\text{O}$ record has been interpreted as a proxy of Antarctic ice volume and oceanic temperature changes across the Oligocene/Miocene boundary^{9,10}. Oscillations in the size of the East Antarctic ice sheet inferred from the ice-proximal CRP-2/2A core show a close correspondence with variations in the $\delta^{18}\text{O}$ record. Two age models have been proposed for the site 929 record. One is based on a nannofossil and planktonic foraminiferal biostratigraphy and, like our CRP-2/2A core data, is calibrated to the MPTS of ref. 33. Durations for intervals between biostratigraphic datums were estimated from inferred ~ 40 -kyr frequencies in magnetic susceptibility records. The second age model is based on a new astronomically calibrated timescale³⁴ established by tuning the 40-kyr cycles in magnetic susceptibility records to orbital target solutions for obliquity and precession³¹. The astronomical chronology has subsequently been used to recalibrate the MPTS in the vicinity of the Oligocene/Miocene boundary³⁶, and indicates that this boundary is ~ 900 kyr younger than that reported in ref. 33. Because both age models for site 929 are based on a fundamental 40-kyr frequency in the magnetic susceptibility data, the durations of geomagnetic chrons and the frequency of the oxygen isotope signal is not affected by changes to absolute ages of the MPTS.

The erosional hiatus of some 250 kyr between sequences 10 and 9 spans an interval of major global cooling, ice-sheet expansion, and sea-level fall culminating at the Mi1 maximum. The oscillation in ice-margin proximity inferred from sequence 9 can be correlated with a ~ 125 -kyr, $\delta^{18}\text{O}$ cycle in the heavily eccentricity-modulated part of the record that follows the Mi1 excursion.

We conclude that the cyclicity recorded by glacial sequences in CRP-2/2A provides strong ice-marginal evidence for orbital oscillations in the size of the Oligocene East Antarctic ice sheet, a style of behaviour identical with that of the unstable Northern Hemisphere ice sheets of the past 2.5 Myr. Studies of Antarctic ice sheets during the Oligocene—a time when planetary temperature was around 3 – 4°C warmer than today—should help to provide realistic analogues for their future behaviour following the increased levels of atmospheric CO_2 and temperature projected for the end of this century³². □

Received 16 November 2000; accepted 3 September 2001.

- Freeman, K. H. & Hayes, J. M. Fractionation of carbon isotopes by phytoplankton and estimates of ancient CO_2 levels. *Glob. Biogeochem. Cycles* **6**, 185–198 (1992).
- Barrett, P. J., Elston, D. P., Harwood, D. M., McKelvey, B. C. & Webb, P.-N. Mid-Cenozoic record of glaciation and sea-level change on the margin of the Victoria Land basin, Antarctica. *Geology* **15**, 634–637 (1987).
- Barrett, P. J. (ed.) Antarctic Cenozoic history from CIROS-1 drillhole, McMurdo Sound. *DSIR Bull.* **245**, 1–254 (1989).
- Hambrey, M. J., Ehrmann, W. U. & Larsen, B. Cenozoic glacial record of the Prydz Bay continental shelf, East Antarctica. *Proc. ODP Sci. Res.* **119**, 77–132 (1991).
- Miller, K. G., Wright, J. D. & Fairbanks, R. G. Unlocking the ice house: Oligocene oxygen isotopes, eustasy and margin erosion. *J. Geophys. Res.* **96**, 6829–6848 (1991).
- Wise, S. W., Breza, J. R., Harwood, D. M. & Wei, W. in *Controversies in Modern Geology* (eds Mueller, D., McKenzie, J. & Weissart, H.) 133–177 (Academic, San Diego, 1991).
- Zachos, J. C., Breza, J. & Wise, S. W. Early Oligocene ice-sheet expansion on Antarctica: stable isotope and sedimentological evidence from Kerguelen Plateau, southern Indian Ocean. *Geology* **20**, 569–573 (1992).
- Zachos, J. C., Quinn, T. M. & Salamy, K. A. High-resolution deep-sea foraminiferal stable isotope records of the Eocene-Oligocene climate transition. *Paleoceanography* **11**, 256–266 (1996).
- Zachos, J. C., Flower, B. P. & Paul, H. Orbitally paced climate oscillations across the Oligocene/Miocene boundary. *Nature* **388**, 567–570 (1997).
- Paul, H. A., Zachos, J. C., Flower, B. P. & Tripathi, A. Orbitally induced climate and geochemical variability across the Oligocene/Miocene boundary. *Paleoceanography* **15**, 471–485 (2000).
- Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
- Hays, D. E. *et al.* General synthesis. *Init. Rep. DSDP* **28**, 919–942 (1975).
- Fielding, C. R., Naish, T. R., Woolfe, K. J. & Lavelle, M. A. Facies analysis and sequence stratigraphy of CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 323–338 (2000).
- Cape Roberts Science Team. Studies from the Cape Roberts Project, Ross Sea, Antarctica, Initial report on CRP-2/2A. *Terra Antarctica* **6**, 1–173 (1999).
- Cape Roberts Science Team. Studies from the Cape Roberts Project, Ross Sea, Antarctica, Initial report on CRP-3. *Terra Antarctica* **7**, 1–209 (2000).
- Henrys, S. A. *et al.* Correlation of seismic reflectors with CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 221–230 (2000).
- Powell, R. D. A model for sedimentation by tidewater glaciers. *Ann. Glaciol.* **2**, 129–134 (1981).
- Naish, T. R. & Kamp, P. J. Sequence stratigraphy of 6th order (41 kyr) Pliocene-Pleistocene cyclothem, Wanganui Basin, New Zealand: A case for the regressive systems tract. *Geol. Soc. Am. Bull.* **109**, 978–999 (1997).
- Passchier, S. Soft sediment deformation features in core from CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 401–412 (2000).
- Van der Meer, J. J. M. Microscopic observations on the first 300 metres of CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 339–348 (2000).
- Talarico, F., Sandroni, S., Fielding, C. F. & Atkins, C. Variability, petrography and provenance of basement clasts in core from CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 529–544 (2000).
- Scherer, R., Bohaty, S. & Harwood, D. Oligocene and lower Miocene siliceous microfossil biostratigraphy for CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 417–442 (2000).
- Bücker, C., Wonik, T. & Jarrard, R. Analysis of downhole logging data from CRP-2/2A, Victoria Land Basin, Antarctica: a multivariate approach. Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 299–310 (2000).
- McIntosh, W. C. ⁴⁰Ar/³⁹Ar geochronology of tephra and volcanic clasts in CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 621–630 (2000).
- Watkins, D. K. & Villa, G. Paleogene calcareous nannofossils from CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 443–452 (2000).
- Lavelle, M. Strontium isotope stratigraphy and age model for CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 611–620 (2000).
- Wilson, G. S., Florindo, F., Sagnotti, L., Roberts, A. & Verosub, K. Magnetostratigraphy of CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 631–646 (2000).
- Wilson, G. S. *et al.* Chronostratigraphy of the CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarctica* **7**, 647–664 (2000).
- Harwood, D. M. & Maruyama, T. Middle Eocene to Pleistocene diatom biostratigraphy of Southern Ocean sediments from the Kerguelen Plateau, Leg 120. *Proc. ODP Sci. Res.* **120**, 683–733 (1992).

30. Hsü, K. J., Percival, S. F., Wright, R. & Peterson, N. P. Numerical ages of magnetostratigraphically calibrated biostratigraphic zones. *Init. Rep. DSDP* 73, 625–635 (1983).
31. Imbrie, J. *et al.* On the structure and origin of major glaciation cycles 2. The 100,000 year cycle. *Paleoceanography* 8, 699–735 (1993).
32. Houghton, J. *et al.* *Climate Change 2001: The Scientific Basis* (Third Assessment Report from IPCC Working Group I), 1–944 (Cambridge Univ. Press, Cambridge, 2001).
33. Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity time scale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* 100, 6093–6095 (1995).
34. Shackleton, N. J., Crowhurst, S. J., Weedon, G. P. & Laskar, J. Astronomical calibration of Oligocene–Miocene time. *Phil. Trans. R. Soc. Lond. A* 357, 1907–1929 (1999).
35. Laskar, J., Joutel, F. & Boudin, F. Orbital, precessional and insolation quantities for the Earth from –20 Myr to +10 Myr. *Astron. Astrophys.* 270, 522–533 (1993).
36. Shackleton, N. J., Hall, M. A., Raffi, I., Tauxe, L. & Zachos, J. Astronomical calibration for the Oligocene–Miocene boundary. *Geology* 28, 447–450 (2000).

Acknowledgements

This Letter is dedicated to the memory of Ken Woolfe, sedimentologist and ISC member of the Cape Roberts Project. The Cape Roberts Project was supported by the Antarctic programmes of Italy, New Zealand, the USA, Germany, Australia, the UK and The Netherlands, with field operations organised by Antarctica New Zealand. We acknowledge the efforts of the Cape Roberts Project International Steering Committee, the Operations/Logistics Management Group, as well as the drilling, logistic support and science teams who provided the material on which this Letter is based. We also acknowledge the support of our home institutions and funding agencies.

Correspondence and requests for materials should be addressed to T.N.
(e-mail: t.naish@gns.cri.nz).

Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes

David Kleijn, Frank Berendse, Ruben Smit & Niels Gilissen

Nature Conservation and Plant Ecology Group, Wageningen University,
Bornsesteeg 69, 6708 PD Wageningen, The Netherlands

Roughly 20% of the European Union's farmland is under some form of agri-environment scheme to counteract the negative impacts of modern agriculture on the environment¹. The associated costs represent about 4% (1.7 billion euros) of the European Union's total expenditure on the Common Agricultural Policy and are expected to rise to 10% in the near future². Although agri-environment schemes have been implemented in various countries for well over a decade, to date no reliable, sufficiently replicated studies have been performed to test whether such measures have the presumed positive effects on biodiversity^{1,3,4}. Here we present the results of a study evaluating the contribution of agri-environment schemes to the protection of biodiversity in intensively used Dutch agricultural landscapes. We surveyed plants, birds, hover flies and bees on 78 paired fields that either had agri-environment schemes in the form of management agreements or were managed conventionally. Management agreements were not effective in protecting the species richness of the investigated species groups: no positive effects on plant and bird species diversity were found. The four most common wader species were observed even less frequently on fields with management agreements. By contrast, hover flies and bees showed modest increases in species richness on fields with management agreements. Our results indicate that there is a pressing need for a scientifically sound evaluation of agri-environment schemes.

Agri-environment schemes cover a wide range of measures, which differ depending on aim, country or even region, but have in common the basis that farmers are paid to adapt the management on (parts of) their farms to the benefit of biodiversity, environment

or landscape. Farmers participate on a voluntary basis, but once they enter a 'management agreement' they are obliged to adhere to a specified set of management prescriptions. We have investigated how effectively the most common form of agri-environment scheme—the management agreement⁵—conserves biodiversity on farms in The Netherlands.

The Netherlands has been implementing this type of agri-environment scheme since 1981, which is considerably longer than comparable European-Union-based measures (EEC regulation 2078/92) that have been introduced after 1992. Dutch agricultural landscapes are particularly important with respect to meadow birds in general, and the wader species black-tailed godwit (*Limosa limosa*) and oystercatcher (*Haematopus ostralegus*) in particular. Roughly 50% and 30–40%, respectively, of the European populations of these two species breed in The Netherlands⁶. As a consequence, most management agreements are aimed to support wader populations. Such agreements oblige farmers to postpone agricultural activities on individual fields until a set date, in June or July, allowing the birds to safely hatch their chicks.

A second type of management agreement is aimed at the conservation of species-rich vegetation in (edges of) grasslands and generally restricts the use of fertilizer and/or postpones the first mowing or grazing date. To evaluate their effectiveness, we used a pair-wise comparison of grassland fields that had long-running management agreements (on average 6 years) with nearby conventionally managed fields (for details see Supplementary Information).

Management agreements designed to enhance the botanical diversity of entire fields or field edges did not have any positive effect on vascular plant species richness in field edges (Fig. 1). No significant differences were found in the composition of the vegetation in edges of fields with or without management agreements. Furthermore, we did not find any correlation between age of the agreement and effect (the difference between each pair of control field and field with management agreement; data not shown). In 31,000 m² of field edge, we found 268, mostly very common, vascular plant species. A comparison between plant species composition of centre and edge (outer 2 m) of the fields showed that plant diversity in agriculturally used Dutch grasslands is a marginal matter indeed. On average, edges accounted for 96% of the total species richness of a field, and 66% of encountered species were never found in the field centre.

Interviews with farmers revealed that nitrogen inputs on fields with management agreements were lower than on conventionally managed fields (106 versus 246 kg N ha^{−1} yr^{−1}; *t*-test, *t*₃₄ = −2.62, *P* = 0.013). Despite this considerable reduction in fertilizer inputs, the current level of inputs, in combination with nitrogen deposition rates of 35–55 kg N ha^{−1} yr^{−1} (ref. 7), may still be too high to promote the development of more species-rich vegetation or to enable less-competitive plant species to establish. In addition, seed

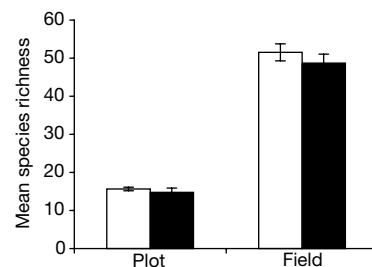


Figure 1 Effects of management agreements aimed at enhancing botanical diversity on number of plant species (mean $\pm$ s.e., *n* = 22) at plot (20 m²) and at field scale (400 m²; sum of 20 plots per field). Open columns, fields with management agreements; filled columns, conventionally managed fields.

sources in intensively used agricultural landscapes are scarce, so even if site conditions are favourable to the establishment of new plant species, dispersal limitations may prevent actual establishment⁸.

There were no positive effects of management agreements on the number of territories or the diversity of all bird or wader species that were observed to nest in the area (Fig. 2a, b). The lack of any positive effect of management agreements on densities or species numbers was independent of the spatial scale that was applied (Fig. 2a, b). Notably, fields with management agreements that aimed to support wader populations by postponing first farming activities had lower counts of the target species lapwing, oystercatcher, common redshank and black-tailed godwit (Table 1). Differences in sward height did not explain the observed preference of waders, because fields with management agreements were avoided both at the start of the growing season (March) and later in the year (June). Many previous studies have shown the selectivity of foraging farmland birds with respect to the management on individual fields^{9,10}. Our study, however, was executed during the breeding season and most observed waders showed territorial behaviour. Lower counts therefore indicate an aversion to use these fields as nest sites, which is highlighted by the significantly fewer oystercatcher territories on fields with management agreements.

Postponing the first mowing or grazing date forced farmers to reduce fertilizer inputs (management agreement compared with conventional: 96 versus 277 kg N ha⁻¹ yr⁻¹; $t_{30} = -3.72$, $P < 0.001$), which probably adversely affected the abundance of the soil animals that waders use for food¹¹. However, management agreements do have a positive effect on reproductive success^{12,13}. Thus, the introduction of management agreements in Dutch agricultural landscapes might have led to an "ecological trap"¹⁴; that is, it might have decoupled the cues that individuals use to select their nesting habitat (for example, food availability) from the main factor that determines their reproductive success (delayed mowing/grazing). Only the starling, an extremely common bird of all lowland habitats including large cities, preferred to forage on fields with management agreements.

Species richness of hover flies was higher on fields with management agreements than on control fields; however, only the initial survey in May contributed significantly to this difference (Fig. 3a;

general linearized model (GLM): $G_{m.a.May} = 15.226$, 1 degree of freedom (d.f.), $P < 0.001$). In contrast to most conventionally managed fields, fields with management agreements had not been mown or grazed in May. Including sward height in the analysis of the May survey showed that the effect of management agreements could be attributed completely to the delayed mowing of these fields, as there were no significant effects of management agreements when fields with similar vegetation heights were compared (Fig. 3b; $G_{m.a.May} = 3.369$, 1 d.f., not significant; $G_{sward\ height.May} = 18.734$, 1 d.f., $P < 0.001$).

Adult hover flies are generalist flower visitors using ubiquitous species from a wide range of families; some hover fly species even feed on grass pollen¹⁵. The observed positive effect of management agreements on the number of hover fly species in the May sample is therefore probably a direct consequence of the delayed mowing date and the resulting prolonged food supply on fields with management agreements.

The bee fauna in Dutch agricultural landscapes is very poor. On average we found only 1.7 species per field and 85% of all caught individuals belong to just three species: the honeybee *Apis mellifera*; and the bumblebees *Bombus pascuorum* and *Bombus terrestris*. But species richness was enhanced by management agreements (Fig. 3a; GLM: $G_{m.a.June} = 14.608$, 1 d.f., $P < 0.001$; $G_{m.a.July} = 4.097$, 1 d.f., $P < 0.05$). For bees, sward height was important but could not explain the observed positive effect (Fig. 3b; $G_{m.a.June} = 8.716$, 1 d.f., $P < 0.01$; $G_{sward\ height.June} = 11.396$, 1 d.f., $P < 0.001$). Apparently, bees perceived a qualitative difference between the two field types that was not shown by the other species groups that we studied.

Our results point out that management prescriptions that have proved to be effective under experimental conditions⁸ do not have the desired effects (plants) or have even unexpected adverse side-effects (birds) when implemented on farms. The motivation and expertise of the farmers may have a crucial role. The primary concern of farmers is necessarily to secure an income. As a result, nature conservation will be of secondary importance to them, and will be fitted into a farming system that, owing to economic pressure, is still increasing in intensity^{16,17}.

Most farmers lack the knowledge to judge in what way measures taken to improve the economic position of their farm (such as grass silaging, lowering the groundwater table) may interfere with nature

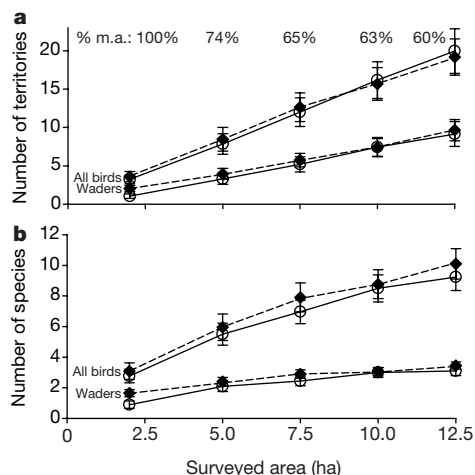


Figure 2 Effects of management agreements on breeding birds at a range of spatial scales. Mean field size 2.0 ha; $n = 20$. % m.a., mean proportion of the management agreement plots actually covered by management agreements (this proportion is 0 at all scales in the control plots). **a**, Density of breeding bird territories; **b**, species richness. Open circles, fields with management agreements; filled diamonds, conventionally managed fields.

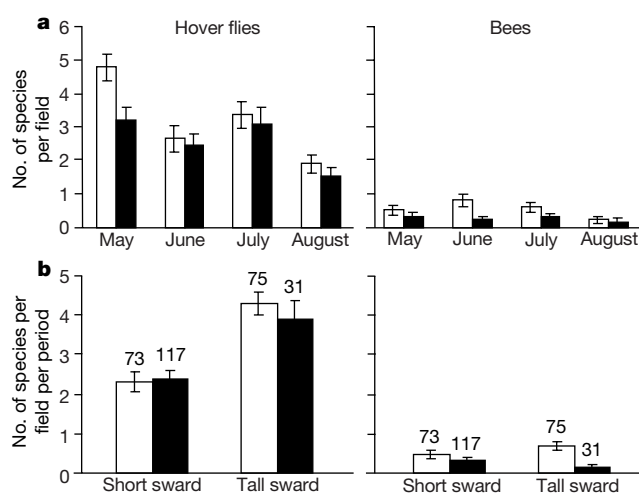


Figure 3 Effects of management agreements on species richness of hover flies and bees. **a**, Mean number of species not previously found on a field. **b**, Interaction between effects of sward height and management agreement (short sward less than ~10 cm, tall sward more than ~10 cm). Open column, fields with management agreements; filled column, conventionally managed fields. $n = 37$. Numbers above columns indicate sample size.

Table 1 Effects of management agreements on mean number of observations and territories per field

Bird species	Counts		Territories	
	MA	Control	MA	Control
Mallard (<i>Anas platyrhynchos</i>)	3.17	4.00	0.43	0.39
Meadow pipit (<i>Anthus pratensis</i>)	1.17	1.09	0.48	0.39
Oystercatcher (<i>Haematopus ostralegus</i>)*	1.52§	2.91	0.13†	0.52
Black-tailed godwit (<i>Limosa limosa</i>)*	2.74†	3.74	0.43	0.43
Starling (<i>Sturnus vulgaris</i>)	5.74§	1.52	0	0
Common redshank (<i>Tringa totanus</i>)*	1.44†	2.61	0.39	0.48
Lapwing (<i>Vanellus vanellus</i>)*	1.30§	3.17	0.26	0.61

The most frequently observed bird species are shown. MA, fields with management agreement. Control, conventionally managed fields. Field size 2.0 ± 0.17 ha (mean $\pm$ s.e., $n = 23$).

* Wader species.

† $P < 0.05$.

‡ $P < 0.01$.

§ $P < 0.001$.

conservation measures that are taken concurrently. By contrast, schemes to promote circl buntings (*Emberiza circlus*) in Devon, UK, have been successful, but these were supervised intensively by scientists¹⁸. Agriculture in The Netherlands is extremely intensive, but can be considered representative for large areas in western Europe (particularly Belgium, France and the UK)³. Many researchers from these countries have come-up with suggestions to extend agri-environment schemes for the specific needs of certain species or ecosystems^{10,19,20}. Our findings indicate that it is imperative to evaluate current agri-environment schemes in all participating countries, and to ensure that any new agri-environment scheme is accompanied by a scientifically sound evaluation plan. □

Methods

Sampling protocol

In 2000, we surveyed plants, birds, hover flies and bees, in nine different areas (3 clayey soil, 3 peat and 3 sandy soil areas were selected randomly from all available areas) throughout the Netherlands. In each area we selected 3–7 field pairs (total 78 fields). The two fields within a pair were located within 1 km of each other; they were similar in size, located in similarly structured parts of the landscape, and had the same soil type and groundwater table.

As diverse vegetation in Dutch grasslands (and the associated insect groups) is predominantly limited to the edges and many farmers declined us access to the centre of the field, the sampling of vascular plants, hover flies and bees focused on the field edges. On each field, plant species composition was determined in 20 quadrats of 2×10 m, bordering and parallel to the ditches separating the fields (some very small fields were sampled with fewer quadrats). A single estimate of the species composition of the field centre was made by walking around the field and noting all observed species. Hover flies and bees were sampled simultaneously by collecting all observed specimens in a 1-m wide 'belt', while walking a transect for 15 min at a constant pace along the field edge (see ref. 21). Specimens were subsequently identified in the laboratory. Samples were taken four times from May to August at monthly intervals. Fields within a pair were always sampled on the same day and by the same person.

The effect of management agreements on birds was measured at different spatial scales ranging from the chosen fields (2.0 ± 0.18 ha; mean $\pm$ s.e.) to 12.5-ha plots around the fields. Birds were surveyed during five field visits between 22 March and 20 June, when all contacts with birds were noted on a map. Subsequently, the numbers of birds holding territory on each field were assessed by methods developed by the Breeding Bird Monitoring Project²², a method bearing resemblance to the one used by the Common Bird Census in the UK. The 12.5-ha plots surrounding control fields contained only other conventionally managed fields; however, the 12.5-ha plots surrounding fields with management agreements contained fields with management agreements as well as conventionally managed fields, because it was impossible to find 12.5-ha plots homogeneously covered by management agreements. The 12.5-ha plot maps were subdivided into progressively larger areas by means of ARCVIEW (mean field size, 5, 7.5, 10, 12.5 ha), which allowed us to analyse the effect of management agreements on bird densities at a range of sampling scales.

Statistical analysis

We distinguished between management agreements aimed at waders and those aimed at plant species richness. Thus, analysis of the effects on waders was performed on a subset of field pairs from which pairs with only 'botanical agreements' were removed. Many fields had both 'meadow bird agreements' and 'botanical agreements'; these were included in

both subsets. Effects on the non-target groups, bees and hover flies, were analysed irrespective of the type of agreement. The study had an unbalanced design (unequal number of pairs nested within areas). We therefore used the residual maximum-likelihood method (REML)²³ followed by Wald-tests²⁴, rather than analysis of variance to test for effects of management agreements.

The data of most individual species, as well as some species groups, contained a high number of zero counts. These data were analysed by means of GLM with a logistic link function and assuming a binomial error distribution, followed by a likelihood ratio test (or G-test). The models included the factors area, pair and management agreement, where both area and pair were considered as replications. As effects of management agreements on the insect groups differed markedly per sampling period, analysis of these species groups was performed on individual sample periods. Furthermore, the factor sward height at the time of sampling was included in the model. All proportional data were arcsin transformed before analysis.

Received 27 July; accepted 2 September 2001.

- Working Document 'State of Application of Regulation (EEC) no. 2078/92. Evaluation of Agri-Environment Programmes', VI/7655/98, 9.11.1998 (01 June 01) (http://europa.eu.int/comm/agriculture/envir/programs/evalrep/text_en.pdf) (1998).
- European Community. For a stronger and wider EU. *Agenda 2000*, Vol. 1. (Office for Official Publications of the European Communities, Luxembourg, 1997).
- Donald, P. F., Green, R. E. & Heath, M. F. Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. R. Soc. Lond. B.* **268**, 25–29 (2001).
- Krebs, J. R., Wilson, J. D., Bradbury, R. B. & Siriwardena, G. M. The second silent spring? *Nature* **400**, 611–612 (1999).
- Beintema, A. J., Dunn, E. & Stroud, D. A. in *Farming and Birds in Europe: the Common Agricultural Policy and its Implications for Bird Conservation* (eds Pain, D. J. & Pienkowski, M. W.) 269–296 (Academic, San Diego, 1997).
- Hagemeijer, W. J. M., Blair, M. J., van Turnhout, C., Bekhuis, J. & Bijlsma, R. *EBCC Atlas of European Breeding Birds: their Distribution and Abundance* (Poyser, London, 1997).
- Berendse, F., Aerts, R. & Bobbink, R. in *Landscape Ecology of a Stressed Environment* (eds Vos, C. C. & Opdam, P.) 104–121 (Chapman and Hall, London, 1993).
- Bakker, J. P. & Berendse, F. Constraints in the restoration of ecological diversity in grassland and heathland communities. *Trends Ecol. Evol.* **14**, 63–68 (1999).
- Milom, T. P., Ennis, D. C., Haskell, D. J., Langton, S. D. & McKay, H. V. Design of grassland feeding areas for waders during winter: the relative importance of sward, landscape factors and human disturbance. *Biol. Conserv.* **84**, 119–129 (1998).
- Henderson, I. G., Vickery, J. A. & Fuller, R. J. Summer bird abundance and distribution on set-aside fields on intensive arable farms in England. *Ecography* **23**, 50–59 (2000).
- Standen, V. Production and diversity of enchytraeids, earthworms and plants in fertilized hay meadow plots. *J. Appl. Ecol.* **21**, 293–312 (1984).
- Beintema, A. J. & Müskens, G. J. D. M. Nesting success of birds breeding in Dutch agricultural grasslands. *J. Appl. Ecol.* **24**, 743–758 (1987).
- Schekkerman, H. & Müskens, G. Do Black-tailed Godwits *Limosa limosa* breeding in agricultural grasslands produce sufficient young for a stable population? *Limosa* **73**, 121–134 (2000).
- Kokko, H. & Sutherland, W. J. Ecological traps in changing environments: ecological and evolutionary consequences of a behaviourally mediated Allee effect. *Evol. Ecol. Res.* **31**, 537–551 (2001).
- Branquart, E. & Hemptinne, J. L. Selectivity in the exploitation of floral resources by hoverflies (Diptera: Syrphinae). *Ecography* **23**, 732–742 (2000).
- MacDonald, D. et al. Agricultural abandonment in mountain areas of Europe: Environmental consequences and policy response. *J. Env. Manag.* **59**, 47–69 (2000).
- Bignal, E. M. Using an ecological understanding of farmland to reconcile nature conservation requirements, EU agricultural policy and world trade agreements. *J. Appl. Ecol.* **35**, 949–954 (1998).
- Evans, A. in *Farming and Birds in Europe: the Common Agricultural Policy and its Implications for Bird Conservation* (eds Pain, D. J. & Pienkowski, M. W.) 331–357 (Academic, San Diego, 1997).
- Chamberlain, D. E., Wilson, A. M., Browne, S. J. & Vickery, J. A. Effects of habitat type and management on the abundance of skylarks in the breeding season. *J. Appl. Ecol.* **36**, 856–870 (1999).
- Ovenden, G. N., Swash, A. R. H. & Smallshire, D. Agri-environment schemes and their contribution to the conservation of biodiversity in England. *J. Appl. Ecol.* **35**, 955–960 (1998).
- Banaszak, J. Studies on methods of censusing the number of bees (Hymenoptera, Apoidea). *Pol. Ecol. Stud.* **9**, 421–505 (1983).
- van Dijk, A. J. *Surveying Breeding Birds in Sample Plots (Manual of the Breeding Bird Monitoring Project)* (in Dutch) (SOVON, Beek-Ubbergen, The Netherlands, 1996).
- Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).
- Committee of the Statistics Department. *Genstat 5 Release 3 Reference Manual* (Oxford Univ. Press, Oxford, 1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank all farmers who gave us permission to work on their fields, and Dienst Landelÿk Gebied for their assistance with the area and field selection. The comments of N. van Breemen, W. ter Keurs, M. Noordervliet and J. Harvey improved this manuscript. Field assistance from B. Brak, J. Smit, M. Gleichman, M. Meijer zu Schlochtern, L. de Nijs and H. Klees is greatly appreciated. This work was funded by the Dutch Stimulation Program Biodiversity.

Correspondence and requests for materials should be addressed to D.K. (e-mail: David.Kleijn@staf.ton.wau.nl).

Endoscopic exploration of Red Sea coral reefs reveals dense populations of cavity-dwelling sponges

Claudio Richter*, Mark Wunsch*, Mohammed Rasheed†‡, Iris Kötter* & Mohammad I. Badran†

* Zentrum für Marine Tropenökologie, Fahrenheitstr. 6, D-28359 Bremen, Germany

† Marine Science Station, University of Jordan and Yarmouk University, PO Box 195, Aqaba, Jordan

‡ Max-Planck-Institut für Marine Mikrobiologie, Celsiusstr., D-28359 Bremen, Germany

Framework cavities are the largest but least explored coral reef habitat¹. Previous dive studies of caverns, spaces below plate corals, rubble and artificial cavities^{1–3} suggest that cavity-dwelling (coelobite) filter-feeders are important in the trophodynamics of reefs^{2,4,5}. Quantitative community data are lacking, however, as the bulk of the narrow crevices interlacing the reef framework are inaccessible to conventional analysis methods⁶. Here we have developed endoscopic techniques to explore Red Sea framework crevices up to 4 m into the carbonate rock, revealing a large internal surface (2.5–7.4 m² per projected m² reef) dominated by encrusting filter-feeders. Sponges alone provided up to 60% of coelobite cover, outweighing epi-reefal filter-feeder biomass by two orders of magnitude. Coelobite community filtration removed more than 60% of the phytoplankton in the course of its less than 5-minute passage through the crevices, corresponding to an uptake of roughly 0.9 g carbon m⁻² d⁻¹. Mineralization of the largely allochthonous organic material is a principal source of nutrients supporting coral and algal growth. The supply of new material by coelobites may provide a key to understanding the ‘coral reef paradox’—a rich ecosystem thriving in nutrient-poor water.

Endoscopic estimates of crevice wall area and coelobite filter-feeder area cover were combined with field data on phytoplankton consumption and mineralization for the first comprehensive assessment of the role of coelobite filter-feeders in the coral reef nutrient balance. Research was carried out as part of the Arab/Israeli/German Red Sea Programme, providing access to various coral reefs in Egypt, Israel and Jordan (Fig. 1). The reefs as well as the characteristics of the study area have been studied in detail by various researchers (refs 5–11; and references therein). They represent typical flourishing Red Sea fringing reefs, characterized by a narrow shelf and a fairly open unconsolidated framework with little sediment infilling.

Line transects showed that 26–42% of the projected reef area is

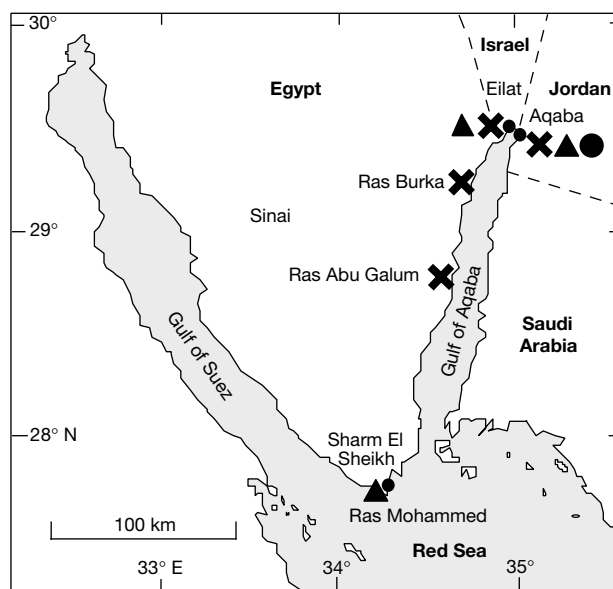


Figure 1 Map of the study area in the northern Red Sea and Gulf of Aqaba with sampling locations for crevice morphology and dimension, coelobite community and water exchange rates (triangles), Chla, phaeopigments and oxygen (crosses), and nutrients (circles).

riddled by crevices of various sizes (Table 1, Fig. 2a). The median opening diameter of only 0.2 m renders these crevices inaccessible to visual inspection by divers using conventional methodology. We carried out detailed measurements of crevice dimensions with a diver-operated endoscopic video-camera (CaveCam⁶) combined with a radially projected light sheet mounted in front of the camera. This arrangement allowed us to outline the shape of the crevices in three dimensions and calculate their wall and cross-sectional areas (Fig. 3a–c; see Methods). Regression analysis showed a linear decrease in the ratio of wall to entrance area with cavity size (Fig. 2b). Combining these results with the line transect data yielded a cumulative coelobite living area of 2.5–7.4 m² of crevice wall per m² reef (Table 1). This is a conservative estimate considering the fact that many of the crevices extended beyond the range of the quantitative survey, and that the interconnections between anastomosing crevices⁷ escaped detection by the light sheet.

We used the CaveCam in a different configuration with small headlights, a 45° mirror device, a close-up lens and spacers⁶ to assess the corresponding community composition and living cover of coelobites (Fig. 3d). Quantitative analysis of 2,301 high-resolution images revealed a rich coelobite community covering 2.8 ± 0.9 m² per projected m² reef, excluding microfacies and sediment-covered areas. Coralline algae predominated near the sunlit entrances. Sponges abounded in posterior sections of the crevices, constituting 51–73% of the coelobite cover (Fig. 4a). The high densities, as well

Table 1 Abundance and characteristics of framework crevice in Red Sea coral reefs

Transect no.	Study site	Transect depth (m)	Transect length (m)	No. of crevices per metre of reef	Metres of crevice per metre of reef	Crevice diameter (m)	s.e.	Maximum diameter (m)	Wall:entrance area	Wall area per projected m ² reef
1	Aqaba	10	20	1.5	0.32	0.21	0.03	0.70	38.8	7.4
2	Aqaba	10	20	0.8	0.26	0.16	0.05	0.60	41.5	5.7
3	Aqaba	14	20	1.45	0.33	0.22	0.04	0.90	38.0	4.9
4	Aqaba	18	20	1.2	0.28	0.23	0.03	0.60	37.7	7.5
5	Ras Mohammed	3	50	1.1	0.42	0.39	0.05	1.80	31.0	4.0
6	Ras Mohammed	12	50	0.94	0.42	0.45	0.08	3.10	29.7	2.5
7	Ras Mohammed	12	50	1.14	0.38	0.33	0.06	2.30	35.0	3.0
8	Ras Mohammed	20	50	0.96	0.40	0.42	0.06	2.10	30.0	3.6
9	Ras Mohammed	20	50	1.14	0.37	0.33	0.04	1.50	33.9	3.8
All	Red Sea	3–20	20–50	1.14	0.35	0.33	0.02	3.10	35.1	4.7

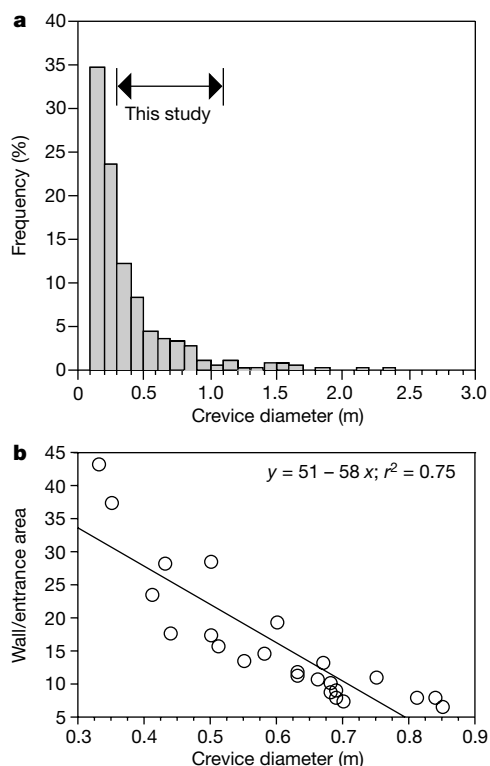


Figure 2 Physical dimensions of coral reef crevices. **a**, Length–frequency histograms of crevice opening diameters, showing the size range amenable to the endoscopic methods used in this study. Earlier studies by divers were limited to cavities with opening diameters much greater than 1 m, comprising less than 1% of the total number of crevices and much less than 1% of the total cavity area. **b**, Surface increase (ratio of crevice wall area to entrance area) as a function of cavity size, highlighting the importance of small crevices as living habitats in the coral reef framework.

as the dominance of delicate sheet-like growth forms (Fig. 3d), support the assumption that the distribution and abundance patterns of coral reef sponges are controlled by predators^{12,13}. Less than 1% of the total area covered by sponges was due to erect or massive morphotypes, and less than 2% was due to boring taxa. Other filter-feeders (ascidians, bivalves, bryozoans and polychaetes) occurred regularly but at much lower densities, covering generally less than 5% of the substrate.

Qualitative wide-angle overviews with the CaveCam mounted on a flexible rod confirmed the pattern of well-flushed and densely populated crevices up to the 4 m reaches of the instrument. With a projected cover of $82 \pm 55\%$ per unit area of coral reef, coelobite

sponges outweighed by far epibenthic sponges ($0.2\text{--}1.2\%$ cover^{8–10}). Using an area:biomass conversion of $25.6\text{ mg C per } 10\text{-cm}^2$ sponge, determined on small fragments of fresh reference material ($r^2 = 0.59$, $n = 25$), this translates into a coelobite sponge biomass of $21.1 \pm 14.2\text{ g C per m}^2$ coral reef (median $\pm$ MAD (median absolute deviation)).

Intense filtering by the coelobite community resulted in marked depletions of phytoplankton chlorophyll *a* (Chl*a*) towards the inner reaches of the crevices ($64 \pm 8\%$ of the freestream waters; Fig. 4b; see Methods), alongside marked decreases in the ratio of Chl*a* to its degradation product phaeopigment (Fig. 4c). These findings are consistent with earlier measurements of bacteria and naked cell depletion in artificial cavities from the Caribbean². Community respiration led to small but significant reductions in oxygen levels relative to freestream waters ($5 \pm 2\%$, Fig. 4d; Kruskal–Wallis test, $P < 0.0001$), reflecting the net heterotrophic nature of the cavity habitat.

Current speeds, determined by video-tracking of displaced particles and by dissolution of calibrated plaster cubes spaced over the length of the crevice, averaged between 0.9 and 5.5 cm s^{-1} . Wash-out experiments with fluorescent dyes featured half-life periods of only $75 \pm 15\text{ s}$, suggesting complete flushing of cavity waters within a few minutes.

Dye experiments showed that water flow through framework crevices was driven by flow speed differences across the bumpy reef surface, much like pressure-induced air flow through termite mounds, where the intake openings are located in troughs near the base and the exhaust openings in exposed position near the crest¹⁴. As a result, water flow was almost always directed into the crevices, leaving the framework through countless cracks and holes near the elevated parts of the reef.

The largely unidirectional flow pattern allowed us to determine the bulk filtering rate of the coelobite community using the standard flow respirometric approach¹⁵. Flux was calculated from the measured changes in Chl*a* and the rate of water exchange across a unit volume of cavernicolous reef, according to

$$F = \Delta\text{Chl}a \times rpk \quad (1)$$

where F is the amount of phytoplankton carbon filtered per unit volume of cavernicolous reef ($\text{g C per m}^3\text{ reef d}^{-1}$, or $\text{g C per m}^2\text{ reef d}^{-1}$ normalized, for conservancy, to the upper first metre of framework); $\Delta\text{Chl}a$ is the mean concentration difference between upstream and cavity waters ($0.16 \pm 0.01\text{ mg Chl}a\text{ per m}^3\text{ water}$; Table 2); r is a conservative value for the water exchange rate in the crevices (the inverse of the water residence time, as determined by fluorescent tracer experiments; 300 per day); p is a conservative value for the volume fraction of crevices per unit framework ($0.3\text{ m}^3\text{ water per m}^3\text{ reef}$; Table 1) and k is a carbon:Chl*a* conversion factor of $60\text{ g C per g Chl}a$ (ref. 10).

Table 2 Differences in chlorophyll *a*, phytoplankton biomass, total picoplankton and nutrient concentrations between cavity and freestream waters above the reef

	Measured					
	$\Delta\text{Chl}a$ ($\mu\text{g l}^{-1}$)	ΔNH_4^+ (μM)	ΔNO_2^- (μM)	ΔNO_3^- (μM)	ΔTIN (μM)	ΔPO_4^{3-} (μM)
Mean	-0.164	0.312	0.037	0.395	0.744	0.048
s.e.	0.014	0.097	0.003	0.043	0.116	0.008
<i>n</i>	32	64	64	64	64	64
<i>P</i>	<0.0001	0.1034	<0.0001	<0.0001	<0.0001	<0.0001
	Calculated					
	Phytoplankton ($\mu\text{g C l}^{-1}$)	Picoplankton ($\mu\text{g C l}^{-1}$)	New TIN (μM)	New PO_4^{3-} (μM)	New TIN (% of measured change)	New PO_4^{3-} (% of measured change)
Mean	-9.84	-19.68	0.248	0.015	33.3	32.2

Picoplankton-derived new nutrients were calculated, assuming a conservative 1:1 biomass ratio between phytoplankton and other picoplankton (such as bacteria)¹⁶ and stoichiometric conversion according to the Redfield ratio. Positive values denote enrichment, negative values denote depletion, relative to the freestream reference 2 m away from the reef. TIN, total inorganic nitrogen.

Phytoplankton uptake by the coelobite community amounted to $0.89 \pm 0.05 \text{ g C m}^{-2} \text{ d}^{-1}$, equivalent to 22% of the gross community metabolism of the entire reef¹⁵. Total picoplankton removal, as suggested by the available biomass of bacteria in tropical waters¹⁶, is probably more than twice this value, ranking our findings among the highest rates reported so far for marine and freshwater sponge

communities^{17,18}. This is corroborated by combining our biomass data ($21.1 \text{ g sponge C m}^{-2}$) with reported food rations in benthic filter-feeders (2–10% body C d^{-1} ; ref. 16), which yields similar values ($0.4\text{--}2.1 \text{ g C m}^{-2} \text{ d}^{-1}$).

Owing to the long doubling times of phytoplankton and bacteria (6–24 h; ref. 19) relative to the residence time of water over the narrow shelf (1–5 h; refs 5, 10), most of the picoplankton consumed in the reef originates from offshore, thus constituting a source of new material for the reef ecosystem.

Nutrient enrichments in the cavities suggest intense mineralization of the organic matter by the crevice biota (Table 2). Nutrient ratios near the Redfield ratio ($\text{N:P} = 15.5$; Table 2) reflect the

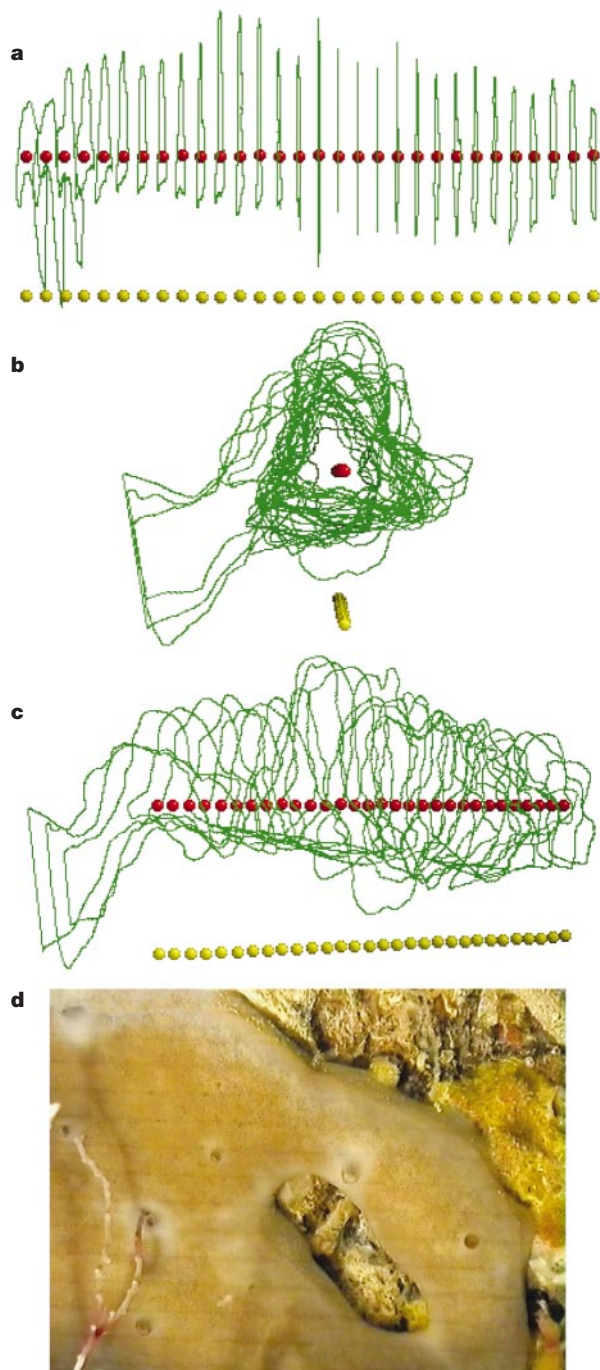


Figure 3 Endoscopic techniques for the study of crevice dimensions and coelobite communities. Wire-frame model of a framework crevice (Ras Mohammed, Egypt, 20 m depth) viewed from the side (a), front (b) and a 45° angle (c). Green circles, spaced 5 cm apart, mark reflection of a light sheet on crevice wall; yellow symbols mark the plumb line. d, Video close-up of coelobite community, including the beige sponge *Chondrilla sacciformis*, an unidentified yellow sponge (at right), solitary scleractinian polyps, the octocoral *Acabaria delicata* (below, left) and polychaete tubes (above, right). Position of the image in d is denoted by a square symbol in a–c.

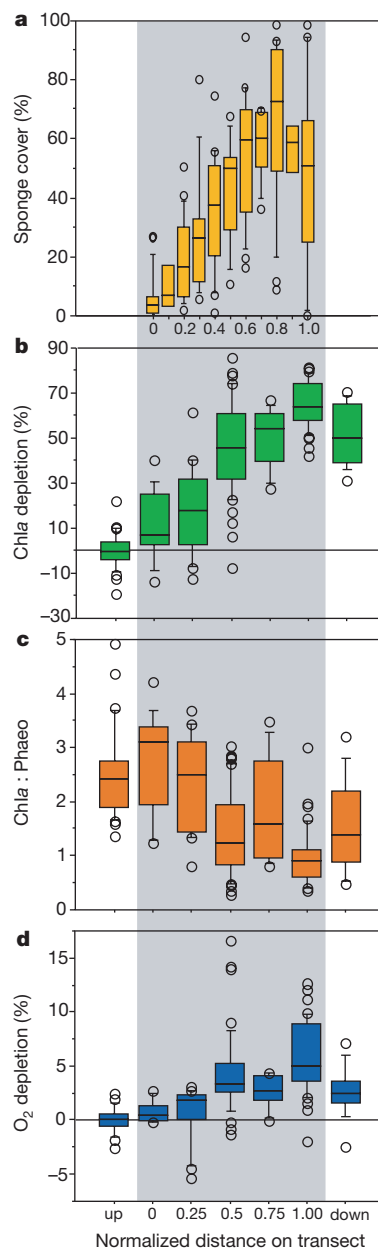


Figure 4 Small-scale distribution of coelobite sponges (a), Chla (b), Chla:phaeopigments (c) and oxygen (d) in Red Sea coral reef framework crevices, shown as composites of 25 (a) and 15 (b–d) surveys conducted within the study area (Fig. 1). Boxes and whiskers encompass 50% and 95% of the data, respectively; centre lines denote the median. In a, per cent cover is relative to total coelobite living area ($2.8 \pm 0.9 \text{ m}^2$ per projected m^2 reef). In b, d, depletions are relative to freestream waters (up) about 2 m above the reef. Downstream exits (down) of tunnel crevices show mixing with freestream waters.

planktonic source of the mineralized material¹⁶, contrasting the higher values reported for intrinsic reef material, for example in lagoonal patch reefs (N:P = 20; ref. 20), pore waters (N:P = 21; ref. 21) or benthic producers (N:P = 30; ref. 22). Stoichiometric conversion of picoplanktonic organic matter to inorganic nutrients (assuming 100% of the ingested food is respired) shows that allochthonous N and P may contribute one-third of the total nutrient flux emanating from the cavities (Table 2), in readily assimilable form (such as ammonia, 42% of N; Table 2) for corals and algae¹⁶.

On the basis of the measured concentration differences and flushing rates, we estimate that 22.3 and 1.4 mmol m⁻² d⁻¹ of allochthonous N and P, respectively, are channelled into the coral reef system by coelobite filter-feeders, which exceeds the known import pathways through cross-shore advection of dissolved nutrients (1.9 and 0.3 mmol m⁻² d⁻¹, respectively²³), nitrogen fixation (0.6–1.0 mmol N m⁻² d⁻¹; ref. 24) or migrating fish (2.4–7.2 mmol N m⁻² d⁻¹; ref. 25).

The accrual of picoplankton by coelobite sponges and the associated enrichment of crevice waters with offshore nutrients may be a widespread phenomenon, as suggested by the occurrence of phyto- and bacterioplankton depletions near coral reefs throughout the tropics^{4,5,10,26,27}. Our findings may therefore provide a general answer to Darwin's question²⁸ of how coral reefs manage to thrive in oligotrophic waters. □

Methods

Crevice numbers and sizes

We performed dive surveys to determine the total number and size distribution of crevices riddling the coral reef framework in Aqaba and Ras Mohammed (Fig. 1). Measuring tapes (50 m) were laid out at random, parallel to the 3-, 10-, 12- and 20-m depth lines (Table 1). Numbers and lengths of crevices intercepting the tape were recorded to the nearest 0.1 m.

Crevice morphology and dimensions

An endoscopic video system was used to assess the cross-sectional and wall area of 25 framework crevices in Aqaba, Eilat and Ras Mohammed (Fig. 1), at depths of 2–5 m ($n = 9$), 12–14 m ($n = 8$) and 19–20 m ($n = 8$). The system consisted of two parts: a camera head fitted with a 3-mm wide-angle lens, connected by a 3.8-m cable to its control (Panasonic KS-162) and video recording unit (Sony TRV-91E)⁶; and a modified 50-W halogen light mounted 60 cm in front of the lens, emitting a plane of light perpendicular to the axis of the camera. This configuration produced a highlighted contour at the intersection of the light sheet with the crevice wall. Moving the set-up in known increments (5 or 10 cm) on a rail along the axis of each crevice yielded a sequence of light rings outlining its shape in three dimensions (Fig. 3a–c). Video-images were digitized, and wall and cross-sectional areas were determined from the stack of scaled images for each crevice using Object-Image 1.62 software written by N. Vischer (ftp://simon.bio.uva.nl/pub). After correction of barrel distortion using Panorama Tools 1.7.2 by H. Dersch (http://www.fh-furtwangen.de/~dersch), three dimensional wire-frame models of the crevices were obtained for visualization (Fig. 3a–c) using Rotater 3.5 by C. Kloeden (ftp://raru.adelaide.edu.au/rotater/).

From the frontal aspect of a given framework crevice, it is obvious that the projected cross-section (Fig. 3b, white area around centre) is only a fraction of the cross-sectional area at the entrance. Given the limited air time underwater, it was not possible to customize the straight track of our system to the winding axis of each crevice, which limits the operational range of the quantitative surveys to 2.5 m. For consistency, the same margin was also applied to the quantitative investigation of the coelobite communities (below).

Coelobite cover

The CaveCam was used with a 7.5-mm close-up lens, 20-W headlights, a 45° mirror and spacers⁶ to assess the corresponding community composition and living cover of coelobites. The walls of each of the 25 crevices were probed in 25-cm increments, taking sets of five 60 × 45-mm video frames of the sides, roof and bottom, respectively. The images were digitized and scaled, and the area covered by each taxon outlined manually with a digitizing pen for image analysis (NIH-Image; http://rsb.info.nih.gov/nih-image). Specimens were determined to the lowest taxonomic level possible and ground-truthed by taxonomic experts on the basis of reference material collected in the field.

Sponge biomass

Sponge material was obtained from small fragments of rock chiselled off the crevice walls. Tissue was scraped off the substrate using a dissecting knife. We obtained 25 samples of coelobite sponges ranging from 11 to 43 cm² in area cover for gravimetric determination of

dry mass (24 h at 90 °C) and ash-free dry mass (AFDM; 5 h at 450 °C). Organic carbon was calculated using a C : AFDM conversion of 0.5 (ref. 16). Each specimen was photographed *in situ* before extraction to relate area cover (image analysis, above) to sponge biomass.

Currents and flushing

We determined water exchange through framework crevices by the dissolution over 24 h of plaster cards²⁹ spaced along the length of the crevices, and by short-term video-tracking of displaced particles using the CaveCam⁶. Additional dye experiments were carried out by injecting fluorescein into the centre of randomly selected cavities, halfway from the entrance, stirring, and measuring the decay of the fluorescence signal in syringe samples taken 0.5, 1, 2, 4, 8 and 16 min after initiation of the experiment. Regression of the log relative fluorescence versus time (seconds) yielded the relationship $y = 1.866 - 0.004t$ ($r^2 = 0.46$; $n = 240$).

Nutrients, oxygen and phytoplankton pigments

Triplicate samples for nutrient, oxygen and chlorophyll determinations were collected by an eight-channel peristaltic pump (Aqaba), which sampled simultaneously in crevice and freestream waters above the reef over a diel period alongside measurements of water exchange. Alternatively, samples were collected by divers (Fig. 1, other sites) drawing water through 100-μm screened silicone tubing into 100-ml polyethylene syringes. Intakes were spaced along the axis of crevices, and additional samples were collected from the downstream ends of tunnel cavities (Fig. 4, right). Cooled and shaded samples were processed within 2 h of collection. Oxygen was measured by Winkler titration³⁰, and Chla and phaeopigments by fluorometry using the acidification method³⁰ (100-ml sample, 25-mm Whatman GF/F filters, 24 h of dark 90% acetone extraction at 4 °C). Filtrate ammonia, nitrite, nitrate and phosphate were determined spectrophotometrically³⁰.

Received 13 March; accepted 21 August 2001.

- Ginsburg, R. N. in *Perspectives on Coral Reefs* (ed. Barnes, D. J.) 148–153 (Australian Institute of Marine Science, Townsville, 1983).
- Buss, L. W. & Jackson, J. B. C. Planktonic food availability and suspension-feeder abundance: evidence of *in situ* depletion. *J. Exp. Mar. Biol. Ecol.* **49**, 151–161 (1981).
- Kobluk, D. R. Cryptic faunas in reefs: ecology and geologic importance. *Palaios* **3**, 379–390 (1988).
- Gast, G. J., Wiegmann, S., Wieringa, E., van Duyl, F. C. & Bak, R. P. M. Bacteria in coral reef water types: removal of cells, stimulation of growth and mineralization. *Mar. Ecol. Prog. Ser.* **167**, 37–45 (1998).
- Richter, C. & Wunsch, M. Cavity-dwelling suspension feeders in coral reefs - a new link in reef trophodynamics. *Mar. Ecol. Prog. Ser.* **188**, 105–116 (1999).
- Wunsch, M. & Richter, C. The CaveCam—an endoscopic underwater videosystem for the exploration of cryptic habitats. *Mar. Ecol. Prog. Ser.* **169**, 277–282 (1998).
- Jackson, J. B. C., Goreau, T. F. & Hartman, W. D. Recent bryozoan-coraline sponge communities and their paleoecological significance. *Science* **173**, 623–625 (1971).
- Mergner, H. in *Proc. Symp. Coastal Marine Environ. Red Sea, Gulf of Aden and Tropical Western Indian Ocean 39–76* (ALECSO Red Sea and Gulf of Aden Environmental Programme, Jeddah (Saudi Arabia), Khartoum, Sudan, 1980).
- Mergner, H. & Schuhmacher, H. in *Proc. 5th Int. Coral Reef Symp.* Vol. 6 (eds Gabrie, C. & Harmelin-Vivien, M.) 243–248 (Antenne Mus., EPHE, Moorea, 1985).
- Yahel, G. *et al.* Phytoplankton distribution and grazing near coral reefs. *Limnol. Oceanogr.* **43**, 551–563 (1998).
- Erez, J. in *Coral Reefs* (ed. Dubinsky, Z.) 411–418 (Elsevier Science, New York, 1990).
- Pawlik, J. R. Coral reef sponges: Do predatory fishes affect their distribution? *Limnol. Oceanogr.* **43**, 1396–1399 (1998).
- Wulff, J. L. Parrotfish predation on cryptic sponges of Caribbean coral reefs. *Mar. Biol.* **129**, 41–52 (1997).
- Vogel, S. *Life in Moving Fluids—the Physical Biology of Flow* 1–467 (Princeton Univ. Press, Princeton, 1994).
- D'Elia, C. F. in *Concepts of Ecosystem Ecology* (eds Pomeroy, L. R. & Alberts, J. J.) 195–230 (Springer, New York, 1988).
- Sorokin, Y. I. *Coral Reef Ecology* (Springer, Berlin, 1995).
- Pile, A. J., Patterson, M. R., Savarese, M., Chernykh, V. I. & Fialkov, V. A. Trophic effects of sponge feeding within Lake Baikal's littoral zone. 2. Sponge abundance, diet, feeding efficiency, and carbon flux. *Limnol. Oceanogr.* **42**, 178–184 (1997).
- Reiswig, H. M. Particle feeding in natural populations of three marine demosponges. *Biol. Bull.* **141**, 568–591 (1971).
- Ferrier-Pagès, C. & Gattuso, J.-P. Biomass, production and grazing rates of pico- and nanoplankton in coral reef waters (Miyako Island, Japan). *Microb. Ecol.* **35**, 46–57 (1998).
- Andrews, J. C. & Müller, H. Space-time variability of nutrients in a lagoonal patch reef. *Limnol. Oceanogr.* **28**, 215–227 (1983).
- Tribble, G. W., Sansone, E. J., Li, Y.-H., Smith, S. V. & Buddemeier, R. W. in *Proc. 6th Int. Coral Reef Symp.* (eds Choat, J. H. *et al.*) 577–582 (Townsville, 1988).
- Atkinson, M. J. & Smith, S. V. C:N:P ratios of benthic marine plants. *Limnol. Oceanogr.* **28**, 568–574 (1983).
- Rougérie, F. Nature et fonctionnement des atolls des Tuamotu (Polynésie Française). *Oceanol. Acta* **18**, 61–78 (1995).
- Shashar, N., Feldstein, T., Cohen, Y. & Loya, Y. Nitrogen fixation (acetylene reduction) on a coral reef. *Coral Reefs* **13**, 171–174 (1994).
- Meyer, J. L., Schultz, E. T. & Helfman, G. S. Fish schools: an asset to corals. *Science* **220**, 1047–1049 (1983).
- Ayukai, T. Retention of phytoplankton and planktonic microbes on coral reefs within the Great Barrier Reef, Australia. *Coral Reefs* **14**, 141–147 (1995).
- Glynn, P. W. Ecology of a Caribbean coral reef, the *Porites* reef flat biotope. Part II. Plankton community with evidence for depletion. *Mar. Biol.* **22**, 1–22 (1973).

28. Darwin, C. *The Structure and Distribution of Coral Reefs* (Smith, Elder & Company, London, 1842).
 29. Jokiel, P. L. & Morrissey, J. I. Water motion on coral reefs: evaluation of the 'clod card' technique. *Mar. Ecol. Prog. Ser.* **93**, 175–181 (1993).
 30. Parsons, T. R., Maita, Y. & Lalli, C. M. *A Manual of Chemical and Biological Methods for Seawater Analysis* (Pergamon, Oxford, 1984).

Acknowledgements

We thank G. Hempel and the participants of the Red Sea Programme for support; the Egyptian, Israeli and Jordanian authorities for sampling permission; A. Abu-Hilal, the staff of the Aqaba Marine Science Station, G. Yahel, R. Yahel, B. Munkes and E. Saadalla for field and laboratory support; U. Diez, I. and J. Zainer for assistance; K. Fabricius, A. Genin, B. Lazar and G. Yahel for discussions; R. van Soest for sponge determinations; and V. Ittekkot and M. Huettel for improving the manuscript. This study was funded by the German Federal Ministry of Education and Research (BMBF).

Correspondence and requests for materials should be addressed to C.R.
 (e-mail: crichter@uni-bremen.de).

Carrying large fuel loads during sustained bird flight is cheaper than expected

Anders Kvist*, Åke Lindström†‡, Martin Green*, Theunis Piersma†§ & G. Henk Visser‡§

* Department of Animal Ecology, Lund University, Ecology Building, S-22362 Lund, Sweden

† Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, Texel, The Netherlands

‡ Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

§ Centre for Isotope Research, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Birds on migration alternate between consuming fuel stores during flights and accumulating fuel stores during stopovers. The optimal timing and length of flights and stopovers for successful migration depend heavily on the extra metabolic power input (fuel use) required to carry the fuel stores during flight^{1,2}. The effect of large fuel loads on metabolic power input has never been empirically determined. We measured the total metabolic power input of a long-distance migrant, the red knot (*Calidris canutus*), flying for 6 to 10 h in a wind tunnel, using the doubly labelled water technique³. Here we show that total metabolic power input increased with fuel load, but proportionally less than the predicted mechanical power output from the flight muscles. The most likely explanation is that the efficiency with which metabolic power input is converted into mechanical output by the flight muscles increases with fuel load. This will influence current models of bird flight and bird migration. It may also help to explain why some shorebirds, despite the high metabolic power input required to fly, routinely make nonstop flights of 4,000 km longer⁴.

Aerodynamic theory predicts that the mechanical power output the flight muscles must generate, to support the weight of the bird and to overcome the drag of the body and wings, increases with fuel load^{5–7}. The available evidence indicates that the aerodynamic models correctly describe the essential physical processes involved⁸. The proposed parameter values, and hence the generated predictions, are also reasonably realistic⁸. However, mechanical power output accounts for only part of the total metabolic power input of a flying bird^{3–8}. When the flight muscles contract, most of the supplied fuel energy is lost as heat. Heat is also lost when respiration and circulation deliver fuel and oxygen to the flight muscles^{5,6} and in

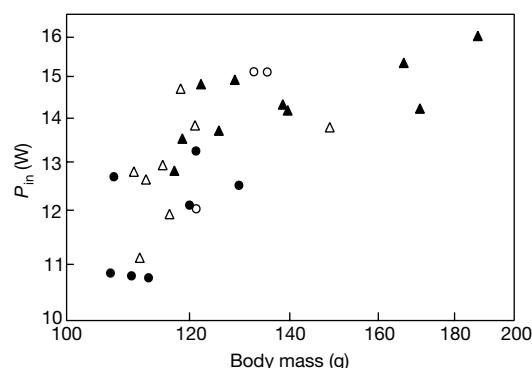


Figure 1 Total metabolic power input (P_{in}) in relation to average body mass (m_b) in red knots flying at 15 m s^{-1} . Different symbols denote different individuals. An analysis of covariance (ANCOVA) with individual as factor and $\log_{10} m_b$ as covariate revealed significant effects of $\log_{10} m_b$ ($F_{1,23} = 7.00$, $P = 0.01$) and individual ($F_{3,23} = 3.11$, $P = 0.046$) on $\log_{10} P_{in}$. The interaction term was not significant (that is, the slope relating $\log_{10} P_{in}$ to $\log_{10} m_b$ did not differ between individuals) and was excluded from the model. The common slope relating $\log_{10} P_{in}$ to $\log_{10} m_b$, correcting for individual differences in intercept, was 0.35 (95% CI 0.08–0.62). Individual intercepts were 0.41, 0.39, 0.41 and 0.35.

metabolic functions not directly involved in the generation of aerodynamic work. In current models of bird flight, metabolic power input is usually calculated from mechanical power output predicted from aerodynamic theory^{5,6,8}. The flight muscles are assumed to convert supplied fuel energy into work with a constant efficiency, the basal metabolic rate is added on top, and a multiplication factor is applied to account for increased respiration and circulation^{5,6,8}. The greatest limitation to these models is the uncertainty of the assumed muscle efficiency⁸.

We measured metabolic power input, using doubly labelled water, in four red knots flying at different body masses at 15 m s^{-1} , in a wind tunnel⁹. The average total metabolic power input during 28 flights lasting between 6 and 10 h was 13.5 W at an average mass of 128 g . The metabolic power input increased allometrically with body mass within individuals with an exponent of 0.35 (95% confidence interval (CI) 0.08–0.62, Fig. 1). Flight muscle efficiency, estimated from metabolic power input and predicted mechanical power output, increased with fuel load (Fig. 2). The estimated level of flight muscle efficiency is uncertain because of uncertainties in the assumptions that come into the calculation of mechanical power^{8,10,11}. However, the significant change in efficiency with fuel load is robust to these uncertainties in the assumptions.

Birds need to carry large fuel stores to perform long migratory flights. The weight of these fuel stores increase the mechanical power output required to fly. Power output from the flight muscles can be controlled by altering the strain rate (wingbeat frequency or amplitude) or the proportion of muscle fibres used^{8,12}. Both of these alternatives affect muscle efficiency^{8,12}. If the flight muscles are optimized to give maximum efficiency at one mechanical power output, efficiency will be reduced as power output changes away from this optimum. Adjustment of flight muscle size and the use of intermittent flight have been suggested to be adaptations for maintaining maximum muscle efficiency as body mass varies owing to build-up and use of fuel stores during migration^{6,12}. Red knots do not use intermittent flight, but they do adjust flight muscle size to their current weight (measured by ultrasound in a parallel study of the same birds)¹³. Great knots (*Calidris tenuirostris*), near relatives of the red knots, reduce flight muscle size during long migratory flights¹⁴. The larger flight muscles at high weights do not seem sufficiently powerful to maintain manoeuvrability, as caged knots are considerably easier to catch when heavy than when light. Several studies have documented how take-off ability in birds is

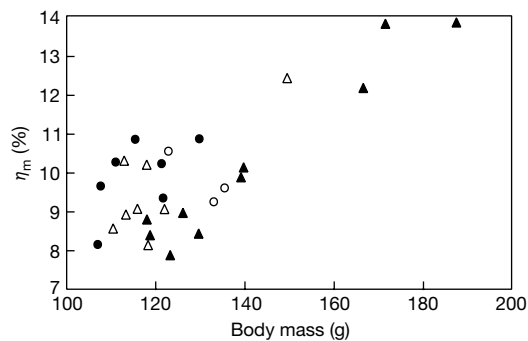


Figure 2 Flight muscle efficiency (η_m) in relation to body mass (m_b). η_m is calculated from measured metabolic power input and mechanical power output predicted from aerodynamic theory (see Methods). An ANCOVA with individual as factor and m_b as covariate showed significant effects of m_b ($F_{1,23} = 71.9$, $P < 0.001$), and individual ($F_{3,23} = 5.86$, $P = 0.004$) on η_m . The interaction term was not significant and was excluded from the model: $\eta_m = \sin(a + 0.0014m_b)^2$, where $a = 0.14, 0.15, 0.13$ and 0.16 for the individual birds. Changes in the assumed induced drag factor, body drag coefficient and profile power ratio within realistic values changes the predicted level of η_m but does not remove the significant effect of m_b on η_m . The η_m values are within the range of other recent studies⁸.

hampered at high body weights^{15,16}. This decrease in manoeuvrability, together with the observed increase in muscle efficiency with body mass, suggests that there may be a trade-off between these two aspects of flight performance. Maintaining flight muscles at a size that gives maximum muscle efficiency in cruising flight may be incompatible with maintaining a large muscle spare capacity for transient manoeuvres, such as sprints, steep climbs and rapid turns. The advantage of maintaining a sizeable muscle spare capacity could be more important when birds are resident in the area and flights are mostly short with a large proportion of transient manoeuvres, for example at stopover sites, winter quarters and breeding quarters. During long migratory flights with heavy fuel loads, selection may instead favour maximal muscle efficiency at the expense of a reduced manoeuvrability. The change in muscle efficiency with body mass may be a result of the birds regulating flight muscle size to maintain an optimal balance between different aspects of flight performance (of which muscle efficiency is one) under changing ecological circumstances. Birds on migratory flights are expected to select an optimal flight speed that changes with body mass¹⁷. During our experiments, birds always flew at one fixed speed. If the flight muscles are adapted to operate efficiently when cruising at an optimal flight speed, part of the variation in muscle efficiency that we observe may be explained by flight speed deviating from the optimal flight speed as body mass changes. An increase in efficiency with body mass implies that the value of a unit of fuel, in potential flight distance, decreases less with fuel load than is assumed in current models of bird migration and flight. The penalty on flying with heavy fuel loads will thus be considerably smaller than previously thought. This significantly alters predictions about the optimal way in which birds organize their migratory journeys and may explain why many shorebirds regularly carry heavy fuel loads and divide their migratory journeys into only a few long nonstop flights measuring 4,000 km or more⁴. □

Methods

Animals and wind tunnel

Four red knots (*Calidris canutus* L.) caught as adults in the Dutch Wadden Sea in 1998 were trained to fly in a wind tunnel⁹. Holding conditions are described in ref. 13. Birds were freely fed trout food pellets supplemented by mealworms (*Tenebrio* sp.) and variation in body mass was due to natural endogenously controlled seasonal changes¹⁸. Safety constructions in the test section included a net upstream, a nylon-covered polyurethane foam ceiling and a net covering the top of the open section. All birds had intact flight feathers and flew in a natural manner, well clear of test section walls during experiments

(Sept.–Dec. 1998, June 1999, Sept.–Dec. 1999 and May 2000). Equivalent air speed was 15 m s^{-1} (within the range for free-flying red knots on migration^{19,20}), turbulence 1% of air speed⁹, air density $1.22\text{--}1.30$ (average: 1.25) kg m^{-3} and temperature $4.7\text{--}14.5$ (average: 10.2) $^{\circ}\text{C}$.

Experiments

At 07:30 local time, after being without food for at least 12 h, the focal bird was injected quantitatively (± 0.001 g, Mettler PM480 DeltaRange), subcutaneously, above the pectoral muscles, with 0.4 ml (10 h flights) or 0.3 ml (6, 7 and 8 h flights) of doubly labelled water (DLW) mixture. Background blood samples were collected before injection. The bird was kept for one hour in a dark chamber ($0.3 \text{ m} \times 0.4 \text{ m} \times 0.3 \text{ m}$) for equilibration of injected isotopes with body water. After this, it was weighed (Mettler PM480 DeltaRange) and a blood sample taken to determine initial enrichment. Flights, lasting 10 h ($n = 17$), 8 h ($n = 1$), 7 h ($n = 1$) or 6 h ($n = 9$), started after a 15-min rest. After the flight the bird was weighed, a blood sample was taken to determine final enrichment and the bird was re-injected with 0.3 ml DLW mixture. After one hour in the small dark chamber, the bird was again weighed and a last blood sample taken to investigate changes in the body water pool during flight on the basis of the principle of isotope dilution²¹. All blood samples were taken by puncturing the brachial vein and collecting 15 μl of blood in each of 5 to 8 microcapillary tubes. The tubes were immediately sealed in a flame and stored at 4°C until analysis. During the first 20 flights birds were weighed after 1, 2, 4, 6 and 8 h, a procedure lasting less than 2 min. On four occasions flights were interrupted for 4 to 21 min because the bird decided to land. Air temperature, air density, interruptions, season or duration of flight did not have a significant effect on the DLW measurements. On average 96% of the time and 99.6% of the energy covered by the DLW measurements was spent in flight. Average mass decrease during experimental flights was 1.09 (standard deviation, $\text{s.d.} = 0.166$) g h^{-1} .

Isotope analysis

Isotope enrichments of blood samples and DLW mixtures for injections were determined in quadruplicate at the Centre for Isotope Research, University of Groningen, The Netherlands²². For every four blood samples, one internal enriched water standard was analysed in quadruplicate to monitor the stability of the infrared mass spectrometer and six gas standard samples for each isotope were used to monitor cross-contamination.

Fractional turnover for ^2H and ^{18}O and CO_2 production were calculated according to equations (3) and (5) in ref. 22, using the first background isotope enrichment for each bird in each season. Dilution spaces were determined separately for ^2H and ^{18}O , before and after the experimental flights, using the plateau method^{13,22,23}. The proportion of water efflux through evaporative pathways is assumed to be 0.85 as determined for a red knot flown under similar circumstances (A.K., personal observations). The calculated CO_2 production is not sensitive to the assumed proportion. A proportion of 0.5 will increase the estimated CO_2 production by 0.8%. Error analyses have revealed that the sensitivity of the DLW method to analytical errors is strongly related to the difference between ^{18}O and ^2H turnover rates³. In birds, the fractional ^{18}O turnover rate is typically about 1.2–1.6 times the ^2H turnover rate³. However, for the red knots flying in the wind tunnel this value was on average 3.18 ($\text{s.d.} = 0.269$, $n = 28$). This phenomenon, in combination with the long flight duration, and quadruplicate isotope analyses, will have resulted in much better precision levels for the experimental birds than usually reported in avian studies. Total metabolic power input (P_{in}) is calculated using an energy equivalent of CO_2 of 14.1 (kJ g^{-1})²⁴ and correcting for energy expended during time not spent flying assuming a resting power of 1.5 times the basal metabolic rate (BMR). Basal metabolic rate was derived from a study of red knots under similar holding conditions²⁵.

Muscle efficiency and the value of fuel

Mechanical power output (P_{out}) is calculated⁶ using the average body mass (m_b , in g) and air density for each flight and the wingspan and aspect ratio (a.r. in m^2) for each bird¹³, assuming an induced drag factor of 1.2 (ref. 6), a body drag coefficient of 0.1 (ref. 10) and a profile power ratio of $(8.4/\text{a.r.})^{11}$. Body frontal area (S_b , in m^2) was taken as $0.00668 m_b^{0.608}$ (average for 22 red knots; B. Spaans, personal communication). Muscle efficiency is calculated as $P_{\text{out}}/(P_{\text{in}}/1.1 - \text{BMR})^6$ and was arcsine transformed before statistical tests.

Received 9 April; accepted 3 September 2001.

1. Alerstam, T. & Lindström, Å. in *Bird Migration: Physiology and Ecophysiology* (ed. Gwinner, E.) 331–351 (Springer, Berlin, 1990).
2. Alerstam, T. & Hedenström, A. The development of bird migration theory. *J. Avian Biol.* **29**, 343–369 (1998).
3. Speakman, J. R. *Doubly Labelled Water: Theory and Practice* (Chapman & Hall, London, 1997).
4. Piersma, T. & Baker, A. J. in *Behaviour and Conservation* (eds Gosling, L. M. & Sutherland, W. J.) 105–124 (Cambridge Univ. Press, Cambridge, 2000).
5. Tucker, V. A. Bird metabolism during flight: evaluation of a theory. *J. Exp. Biol.* **58**, 689–709 (1973).
6. Pennycuik, C. J. in *Avian Biology* Vol. 5 (eds Farner, D. S. & King, J. R.) 1–75 (Academic, London, 1975).
7. Rayner, J. M. V. A new approach to animal flight mechanics. *J. Exp. Biol.* **80**, 17–54 (1979).
8. Rayner, J. M. V. & Ward, S. in *Proc. 22 Int. Ornithol. Congr., Durban* (eds Adams, N. J. & Slotow, R. H.) 1786–1809 (BirdLife South Africa, Johannesburg, 1999).
9. Pennycuik, C. J., Alerstam, T. & Hedenström, A. A new low turbulence wind tunnel for bird flight experiments at Lund University, Sweden. *J. Exp. Biol.* **200**, 1441–1449 (1997).
10. Pennycuik, C. J., Klaassen, M., Kvist, A. & Lindström, Å. Wingbeat frequency and the body drag anomaly: wind-tunnel observations on a thrush nightingale (*Luscinia luscinia*) and a teal (*Anas crecca*). *J. Exp. Biol.* **199**, 2757–2765 (1996).
11. Pennycuik, C. J. The use and misuse of mathematical flight models. *Isr. J. Zool.* **41**, 307–319 (1995).

12. Rayner, J. M. V. Bounding and undulating flight in birds. *J. Theor. Biol.* **117**, 47–77 (1985).
13. Lindström, Å., Kvist, A., Piersma, T., Dekinga, A. & Dietz, M. Avian pectoral muscle size rapidly tracks body mass changes during flight, fasting and fuelling. *J. Exp. Biol.* **203**, 913–919 (2000).
14. Battley, P. F. *et al.* Empirical evidence for differential organ reductions during trans-oceanic bird flight. *Proc. R. Soc. Lond. B* **267**, 191–196 (2000).
15. Metcalfe, N. B. & Ure, S. E. Diurnal variation in flight performance and hence predation risk in small birds. *Proc. R. Soc. Lond. B* **261**, 395–400 (1995).
16. Kullberg, C., Fransson, T. & Jacobsson, S. Impaired predator evasion in fat blackcaps (*Sylvia atricapilla*). *Proc. R. Soc. Lond. B* **263**, 1671–1675 (1996).
17. Hedenström, A. & Ålerstam, T. Optimal flight speeds in birds. *Phil. Trans. R. Soc. Lond. B* **348**, 471–487 (1995).
18. Piersma, T. *et al.* Variability in basal metabolic rate of a long-distance migrant shorebird (red knot, *Calidris canutus*) reflects shifts in organ sizes. *Physiol. Zool.* **69**, 191–217 (1996).
19. Noer, H. Speeds of migrating waders *Charadriidae*. *Dansk Ornitolog. Foren. Tidsskr.* **73**, 215–224 (1979).
20. Gudmundsson, G. A. Spring migration of the knot *Calidris c. canutus* over southern Scandinavia, as recorded by radar. *J. Avian Biol.* **25**, 15–26 (1994).
21. Ancel, A., Visser, G. H., Handrich, Y., Masman, D. & Le Maho, Y. Energy saving in huddling penguins. *Nature* **385**, 304–305 (1997).
22. Visser, G. H., Boon, P. E. & Meijer, H. A. J. Validation of the doubly labelled water method in Japanese Quail *Coturnix c. japonica* chicks: is there an effect of growth rate? *J. Comp. Physiol. B* **170**, 365–372 (2000).
23. Visser, G. H., Dekinga, A., Achterkamp, B. & Piersma, T. Ingested water equilibrates isotopically with the body water pool of a shorebird with unrivalled water fluxes. *Am. J. Physiol.* **209**, R1795–R1804 (2000).
24. Schmidt-Nielsen, K. *Animal Physiology: Adaptation and Environment* 169–173 (Cambridge Univ. Press, Cambridge, 1997).
25. Piersma, T., Cadée, N. & Daan, S. Seasonality in basal metabolic rate and thermal conductance in a long-distance migrant shorebird, the knot (*Calidris canutus*). *J. Comp. Physiol. B* **165**, 37–45 (1995).

Acknowledgements

We thank A. Dekinga, M. W. Dietz, A. Forslid, A. Hedenström, M. Klaassen, A. Koolhaas, M. Rosén and B. Spaans for catching and maintaining birds and participating in experiments; T. Ålerstam and C. J. Pennycuik for their efforts to build the Lund wind tunnel; T.A. for comments; and B. Versteppen for isotope analyses. Thanks to S. Nijens for comments on an earlier version of the manuscript. Supported by grants from the Netherlands Organization for Scientific Research (PIONIER, NWO), the Crafoord Foundation, the Knut and Alice Wallenberg Foundation, the Swedish Council for Planning and Co-ordination of Research and the Swedish Natural Science Research Council. The experiments were carried out under licence from the Lund/Malmö Ethical Committee.

Correspondence and requests for materials should be addressed to A.K. (e-mail: anders.kvist@zoekol.lu.se).

Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3

Lena Alexopoulou*, Agnieszka Czopik Holt†, Ruslan Medzhitov*‡§ & Richard A. Flavell*‡§

* Section of Immunobiology, † Department of Molecular, Cellular and Developmental Biology, ‡ Howard Hughes Medical Institute, and Yale University School of Medicine, New Haven, Connecticut, 06520, USA

§ These authors contributed equally to the work

Toll-like receptors (TLRs) are a family of innate immune-recognition receptors that recognize molecular patterns associated with microbial pathogens, and induce antimicrobial immune responses^{1,2}. Double-stranded RNA (dsRNA) is a molecular pattern associated with viral infection, because it is produced by most viruses at some point during their replication³. Here we show that mammalian TLR3 recognizes dsRNA, and that activation of the receptor induces the activation of NF- κ B and the production of type I interferons (IFNs). TLR3-deficient (TLR3^{-/-}) mice showed reduced responses to polyinosine–polycytidylic acid (poly(I:C)), resistance to the lethal effect of poly(I:C) when sensitized with D-galactosamine (D-GalN), and reduced production of inflammatory cytokines. MyD88 is an adaptor protein that is shared by all the known TLRs¹. When activated by poly(I:C), TLR3 induces cytokine production through a signalling pathway dependent on

MyD88. Moreover, poly(I:C) can induce activation of NF- κ B and mitogen-activated protein (MAP) kinases independently of MyD88, and cause dendritic cells to mature.

Viral infection of mammalian cells results in activation of an innate immune response mediated by type I IFNs, IFN- α and IFN- β , and other cytokines, including interleukin (IL)-6 and IL-12 (refs 4, 5). While IFNs inhibit virus replication, IL-6 and IL-12, which are also induced by bacterial infections, elicit cytotoxic responses needed for elimination of intracellular pathogens. Mammalian TLRs recognize lipopolysaccharide (LPS) and other microbial products^{1,6–10}. Whereas the receptors for LPS are expressed on the cell surface, dsRNA is known to bind only intracellular targets, including the dsRNA-dependent protein kinase (PKR)¹¹. However, cells derived from PKR-deficient (PKR^{-/-}) mice still respond to poly(I:C), a synthetic dsRNA analogue, suggesting the existence of another receptor, which recognizes dsRNA^{12,13}.

To test whether dsRNA can be recognized by a TLR, 293T cells

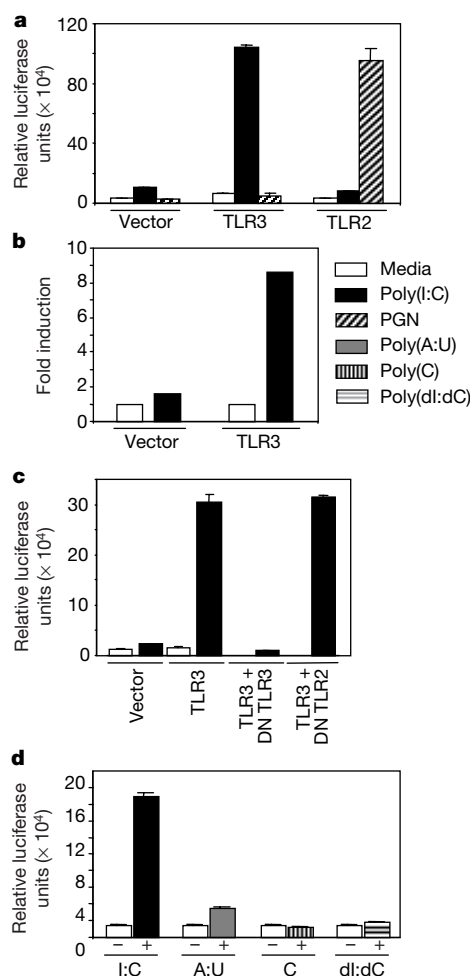


Figure 1 TLR3 specifically signals for NF- κ B activation in response to poly(I:C). **a**, 293T cells were transiently transfected with 50 ng of human TLR3, TLR2 or empty pcDNA3 vector together with an NF- κ B luciferase reporter. Luciferase activity in cells treated with 25 μ g ml⁻¹ poly(I:C) or 10 μ g ml⁻¹ PGN or untreated (media) cells was measured. **b**, Luciferase activity in CaCo-2 cells transiently transfected with 500 ng of empty vector or TLR3 DNA, together with 200 ng NF- κ B luciferase reporter and stimulated with 25 μ g ml⁻¹ poly(I:C). **c**, 293T cells transiently transfected with expression vector for TLR3 or empty vector, together with NF- κ B luciferase reporter and, where indicated, 1 μ g of dominant negative (DN) TLR3 or DN TLR2 DNAs. NF- κ B-induced luciferase activity in cells treated with 25 μ g ml⁻¹ poly(I:C) or untreated cells was measured. **d**, Transfection of RAW 264.7 macrophages with a NF- κ B luciferase reporter. Luciferase activity in cells treated with 20 μ g ml⁻¹ poly(I:C), poly(A:U), poly(C) or poly(dI:dC), or untreated cells.

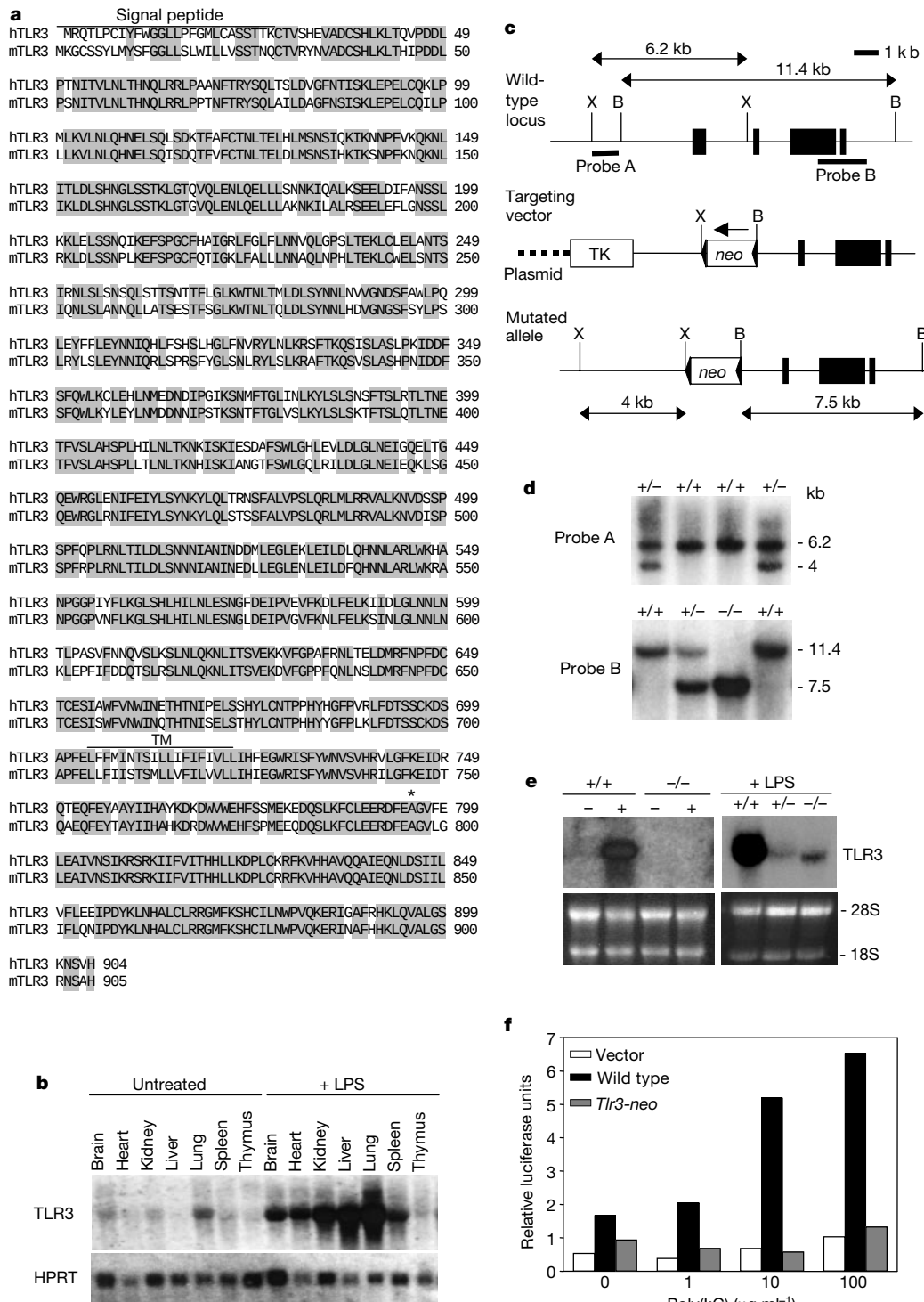


Figure 2 Amino-acid sequence and expression pattern of TLR3, and generation of TLR3^{-/-} mice. **a**, Comparison of the human (h) and murine (m) (GenBank accession number AF355152 and AF40279, respectively) TLR3 sequences. Areas of amino-acid identity (79%) are shaded. The predicted signal peptide and transmembrane domain (TM) are indicated. Asterisk, alanine residue represented by a conserved proline residue in the other TLRs. **b**, Northern blot analysis of murine TLR3. Total RNA extracted from multiple tissues of mice killed 0 or 4 h after intraperitoneal LPS (500 μg) injection was electrophoresed, transferred to a nylon membrane, and hybridized with a 980-bp *Sma*–*Xba*I fragment, containing part of the third exon of *Tlr3* gene as a probe. **c**, The structure of the *Tlr3* gene, the targeting vector and the predicted mutated locus. Filled boxes are exons, filled triangles are *loxP* sites, and open boxes are the selection marker genes. The locations of the 5' external and 3' internal probes are shown. Restriction enzymes: B, *Bam*HI; X, *Xba*I. TK, thymidine kinase gene. **d**, Southern blot analysis of

genomic DNA from embryonic stem-cell clones and from mice digested with *Xba*I and *Bam*HI, respectively. DNA from wild-type (+/+) and targeted embryonic stem (+/-) clones (top), and from wild-type (+/+), heterozygous (+/-) and homozygous (-/-) TLR3 mutant mice (bottom) are indicated. **e**, Northern blot analysis of RNA from thioglycollate-elicited uninduced (-) or LPS-induced (+) macrophages from wild-type (+/+), heterozygous (+/-) and homozygous (-/-) TLR3 mutant mice. Hybridization using the TLR3 N-terminal (left) or C-terminal (right) fragment as a probe is shown. Ethidium bromide staining after RNA transfer to the membrane is included as control (bottom panels). **f**, Wild-type *Tlr3* but not *Tlr3-neo* activates NF- κ B. 293T cells were transiently transfected with wild-type *Tlr3*, *Tlr3-neo* or empty expression vectors together with a NF- κ B luciferase reporter. Luciferase activity is expressed as the ratio of NF- κ B-dependent firefly luciferase activity divided by control *Renilla* luciferase activity. Data are representative of two independent experiments.

that do not respond to stimulation with poly(I:C) (Fig. 1a) were used in an *in vitro* system. We used 293T cells expressing one of a range of TLRs (human TLR1–6 or TLR9) together with an NF- κ B-dependent reporter gene. When stimulated with poly(I:C), only those cells expressing human TLR3 showed marked responsiveness to poly(I:C); the cells expressing the rest of the TLRs, including

TLR2, did not (Fig. 1a and data not shown). To demonstrate the specificity of the ligand, we used peptidoglycan (PGN), which activated TLR2, as previously reported⁹, but not TLR3 (Fig. 1a). Similar results were obtained when we used the human Caco-2 cells, which also do not respond to poly(I:C) (Fig. 1b). Moreover, co-transfection of a dominant negative version of TLR3 in 293T cells,

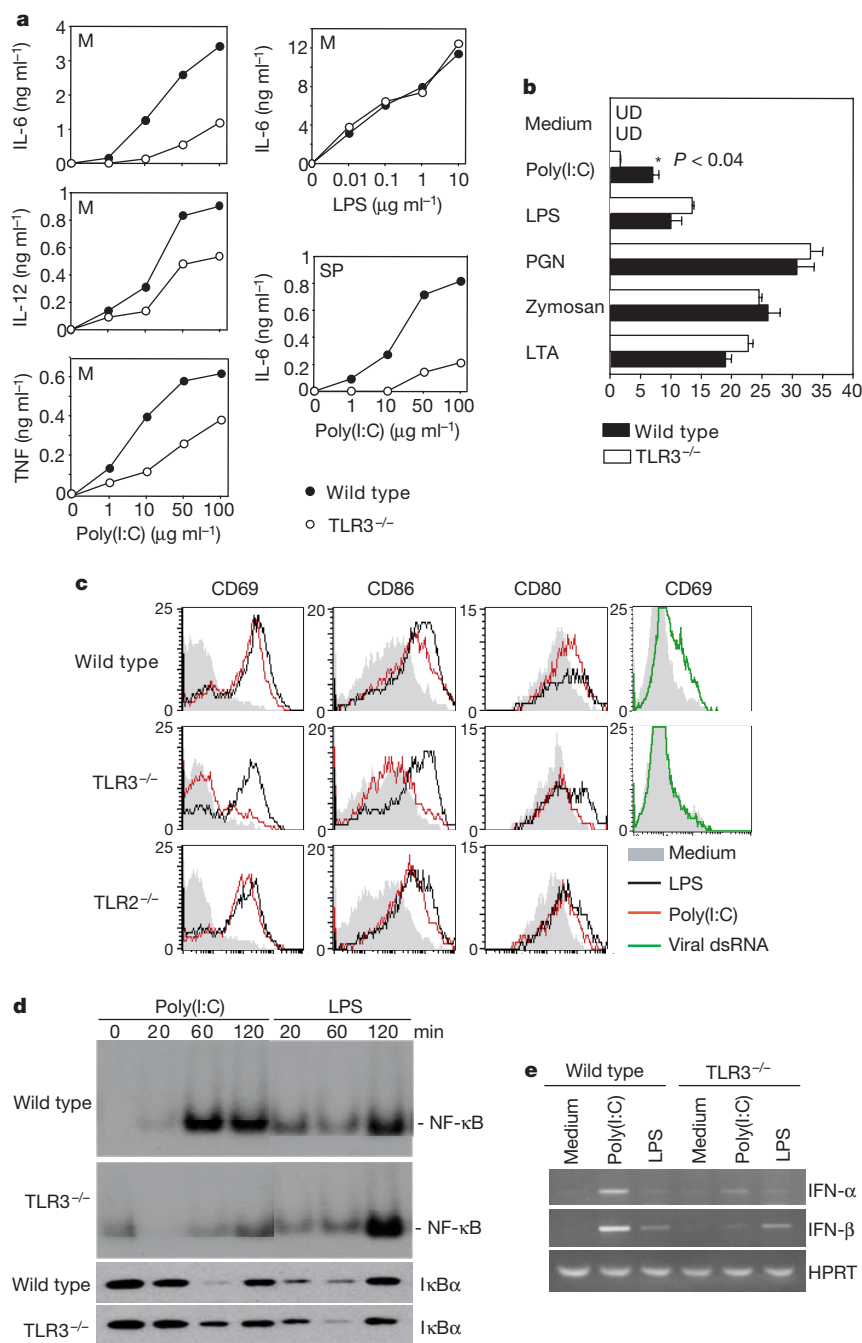


Figure 3 Impaired responses to poly(I:C) from TLR3^{-/-} cells. **a**, Bone-marrow-derived macrophages (M) or total splenocytes (SP) from wild-type or TLR3^{-/-} mice were stimulated with poly(I:C) or LPS for 24 h, and concentrations of IL-6, IL-12 p40/p70 and TNF- α in the culture supernatants were measured by ELISA. Data are representative of three independent experiments. **b**, Bone-marrow-derived macrophages from wild-type or TLR3^{-/-} mice were stimulated with 100 μ g ml⁻¹ poly(I:C), 10 μ g ml⁻¹ LPS, 10 μ g ml⁻¹ PGN, 100 μ g ml⁻¹ zymosan or 10 μ g ml⁻¹ LTA for 24 h, and concentrations of IL-6 in the culture supernatants were measured by ELISA. Data are representative of two independent experiments. UD, undetected. **c**, Splenocytes were cultured for 24 h with

100 μ g ml⁻¹ poly(I:C), 5 ng ml⁻¹ LPS or 30 μ g ml⁻¹ viral dsRNA, or left untreated. Cells were collected, stained and analysed by flow cytometry. Histograms show expression levels of CD69, CD86 and CD80 in B220⁺ (B cells) gated lymphocyte populations. **d**, Bone-marrow-derived macrophages from wild-type or TLR3^{-/-} mice were stimulated with 20 μ g ml⁻¹ poly(I:C) or 5 ng ml⁻¹ LPS. At the indicated points, cells were lysed; nuclear translocation of NF- κ B was visualized by EMSA, and degradation of I κ B α was analysed by western blot. **e**, Bone-marrow-derived macrophages from wild-type or TLR3^{-/-} mice were stimulated with 100 μ g ml⁻¹ poly(I:C) or 1 μ g ml⁻¹ LPS, or left untreated. After 4 h, total RNA was isolated and expression of IFN- α , IFN- β and HPRT was determined by RT-PCR.

but not of a dominant negative TLR2, completely blocked poly(I:C) signalling (Fig. 1c), indicating that TLR3 confers responsiveness to poly(I:C).

Treatment of a RAW 264.7 macrophage cell line with poly(I:C) resulted in activation of the NF- κ B reporter gene (Fig. 1d). Additionally, polyadenylic–polyuridylic acid (poly(A:U)), which is another dsRNA polymer, activated RAW macrophages in a manner similar to poly(I:C), but significantly less strongly (Fig. 1d). Importantly, neither polycytidylic acid (poly(C)), a synthetic single-stranded RNA analogue, nor polydeoxyinosinic–deoxycytidylic acid (poly(dI:dC)) had any effect on NF- κ B activation in RAW cells (Fig. 1d), as expected.

To characterize the biological function of murine TLR3, we generated TLR3^{-/-} mice by homologous recombination in embryonic stem cells. To isolate the murine TLR3 gene, we screened a 129/SvJ mouse genomic library with a 980-base pair (bp) complementary DNA fragment of human TLR3 as a probe. Sequence analysis revealed the presence of regions conserved in the TLR family, such as leucine-rich repeats (LRRs) and a Toll/IL-1 receptor (TIR) homology domain (Fig. 2a). However, the conserved proline residue within the cytoplasmic domain of the murine and human TLR proteins that is mutated in LPS-hyporesponsive C3H/HeJ mice (P712H)¹⁴ is represented by an alanine residue in TLR3 (Fig. 2a). Murine TLR3 transcripts were most abundantly expressed in the lung, brain and kidney as detected by northern blot analysis (Fig. 2b). However, when LPS was injected intraperitoneally, a dramatic upregulation of expression of *TLR3* messenger RNA could be seen in all tissues tested except thymus, suggesting that the expression of TLR3 is inducible (Fig. 2b).

To generate TLR3^{-/-} mice, we replaced the first exon of the *Tlr3* gene with a neomycin-resistance cassette (*neo*) flanked by two *loxP* sites (Fig. 2c). Chimaeric mice were produced by microinjecting embryonic stem cells from three correctly targeted clones into C57BL/6 blastocysts. Male chimaeras were mated to C57BL/6 females and the transmission of the mutated allele through the germ line was confirmed in all three lines by Southern blot analysis (Fig. 2d). Northern blot analysis showed that *Tlr3* transcript from the mutant mice was absent or detected as smaller and in reduced amounts compared with that from the wild-type mice, when an amino-terminal or a carboxy-terminal fragment was used as a probe, respectively (Fig. 2e). Because low levels of a truncated *Tlr3* mRNA were detected, a transient transfection assay was used to assess the functional activities of wild-type (*Tlr3*) and mutant *Tlr3* (*Tlr3-neo*) genes. The genes were cloned into an expression vector and transiently transfected into 293T cells together with an NF- κ B luciferase reporter gene. When the cells were stimulated with poly(I:C), substantial activation of NF- κ B was observed only in the wild-type *Tlr3* cells (Fig. 2f), indicating that the truncated transcript

Tlr3-neo expressed in the mutant mice does not encode a functional protein.

TLR3^{-/-} mice had normal appearance, growth, size, fertility and lifespan, and showed no obvious behavioural abnormalities. Moreover, flow cytometry revealed that the expression of CD3, B220, CD4 and CD8 in thymocytes and splenocytes were not altered in TLR3^{-/-} mice compared with wild-type mice (data not shown).

Our *in vitro* experiments demonstrated that stimulation of TLR3 by poly(I:C) leads (directly or indirectly) to the activation of NF- κ B (Fig. 1). We therefore analysed responses of TLR3^{-/-} cells to poly(I:C). Macrophages from wild-type mice produced IL-6, IL-12 and TNF- α in a dose-dependent manner in response to poly(I:C) (Fig. 3a). However, the ability of TLR3^{-/-} macrophages to produce these inflammatory cytokines in response to poly(I:C) was significantly impaired (Fig. 3a). Moreover, the production of IL-6 from TLR3^{-/-} total splenocytes was impaired compared with that of wild-type cells (Fig. 3a). No significant differences were observed in the production of TNF- α , IL-6 or IL-12 between wild-type and TLR3^{-/-} macrophages in response to LPS, PGN, lipoteichoic acid (LTA), zymosan, mannan or CpG DNA (Fig. 3b and data not shown), indicating that there were no detectable intrinsic defects in cytokine production in TLR3^{-/-} macrophages other than their specific defect in their response to poly(I:C).

To demonstrate the specificity of recognition of poly(I:C) by TLR3, we tested the activation of B cells by TLR3^{-/-} cells and compared it with that by TLR2-deficient (TLR2^{-/-})⁹ and wild-type cells. Total splenocytes from wild-type, TLR2^{-/-} and TLR3^{-/-} mice were stimulated with poly(I:C) or LPS and the expression of CD69, CD80 and CD86 on B cells was analysed by flow cytometry (Fig. 3c). A comparable augmentation of the expression of these surface molecules was observed in wild-type and TLR2^{-/-} B cells (Fig. 3c). In contrast, TLR3^{-/-} B cells did not show any response to poly(I:C), despite exhibiting normal responses to LPS (Fig. 3c), indicating that TLR3 is important for cellular responses to poly(I:C) but not to LPS. Similar results were obtained with macrophages (data not shown). We then tested the responses of TLR3^{-/-} and wild-type splenocytes when stimulated with viral genomic dsRNA from type I Lang mammalian reovirus. In contrast to wild-type B cells, which upregulated CD69, TLR3^{-/-} cells showed no response to viral dsRNA (Fig. 3c), suggesting that TLR3 is important for recognizing not only synthetic poly(I:C), but also viral dsRNA.

Signalling pathways that are responsive to viral infection or dsRNA involve activation of NF- κ B and c-Jun N-terminal kinase (JNK)^{5,12,13}. We next analysed poly(I:C)-induced NF- κ B activation at different time points by electrophoretic mobility shift assay. In wild-type macrophages, binding activity of NF- κ B was detected at 60 min and sustained for up to 120 min. In contrast, in TLR3^{-/-} macrophages, activation of NF- κ B was barely detectable at 60 min,

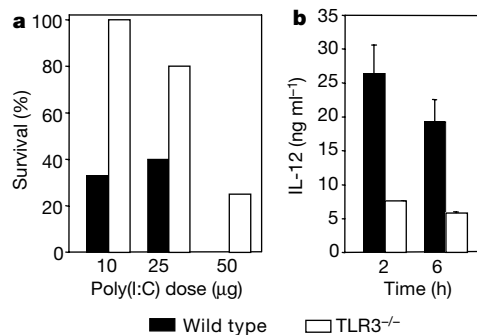


Figure 4 Measurement of poly(I:C)-induced shock in TLR3^{-/-} mice. **a**, Mice (8–12 weeks old) were injected intraperitoneally with the indicated amounts of poly(I:C) and 20 mg D-GalN (Sigma). The dose of poly(I:C) was calculated per 20 g of body mass. Survival was monitored for 3 d. Three, five and four mice were used for 10, 25 and 50 µg poly(I:C),

respectively. **b**, TLR3^{-/-} and wild-type mice were injected intraperitoneally with 50 µg poly(I:C). Sera were taken 2 and 6 h later and serum levels of IL-12 were determined with specific ELISA assays. The data are means $\pm$ s.e. of sera samples from two and four mice at 2 and 6 h, respectively.

and was not increased even after 120 min (Fig. 3d). However, TLR3^{-/-} macrophages stimulated by LPS activated NF- κ B to the same extent as did wild-type cells (Fig. 3d). Analysis of the levels of I κ B α protein in cytoplasmic extracts yielded similar results (Fig. 3d).

Type I IFNs are induced by virtually every type of viral infection, at least in part through the detection of viral dsRNA, and mediate their antiviral activities through the induction of cellular proteins that in turn participate in RNA degradation, inhibition of translation, and regulation of major histocompatibility complex (MHC) antigens⁵. To determine whether TLR3 contributes to induction of type I IFN genes, we used macrophages derived from wild-type and TLR3^{-/-} mice. Little induction of IFN- α or IFN- β mRNA was observed in TLR3^{-/-} macrophages when treated with poly(I:C) compared with wild-type cells (Fig. 3e). When stimulated with LPS, the induction of type I IFNs was much less pronounced, but was similar in wild-type and TLR3^{-/-} macrophages (Fig. 3e). This result indicates that the transcriptional induction by dsRNA of genes for IFN- α and - β is dependent on TLR3.

Next, we investigated whether signalling by TLR3 is required for the *in vivo* inflammatory response to poly(I:C)¹⁵. TLR3^{-/-} and wild-type mice were injected intraperitoneally with poly(I:C) upon D-

GalN sensitization. TLR3^{-/-} mice were significantly resistant to poly(I:C)-induced shock compared with wild-type mice (Fig. 4a). Moreover, when injected with poly(I:C), TLR3^{-/-} mice exhibited reduced production of IL-12 in the sera compared with wild-type controls (Fig. 4b).

Members of the TLR family activate the Rel-family transcription factor NF- κ B through the adapter protein MyD88, the serine/threonine kinase IRAK, and the tumour necrosis factor (TNF)-receptor-associated factor 6 (TRAF6) (refs 1, 16). To test whether MyD88 is involved in dsRNA-induced signalling, we analysed poly(I:C)-induced maturation of dendritic cells and activation of macrophages derived from MyD88-deficient (MyD88^{-/-}) mice¹⁷. Stimulation of wild-type dendritic cells with poly(I:C) induced production of IL-12 (Fig. 5a) and IL-6 (data not shown) in a dose-dependent manner, whereas only background levels were detectable in MyD88^{-/-} dendritic cells (Fig. 5a). Similar results were obtained with macrophages (data not shown). Moreover, MyD88^{-/-} macrophages were also deficient in secreting NO₂ when treated with poly(I:C) or LPS (Fig. 5b). However, denatured poly(I:C) or single-stranded RNA, such as poly(C), did not activate wild-type or MyD88^{-/-} macrophages (Fig. 5b) or dendritic cells (data not shown). Additionally, MyD88^{-/-} total splenocytes showed only a

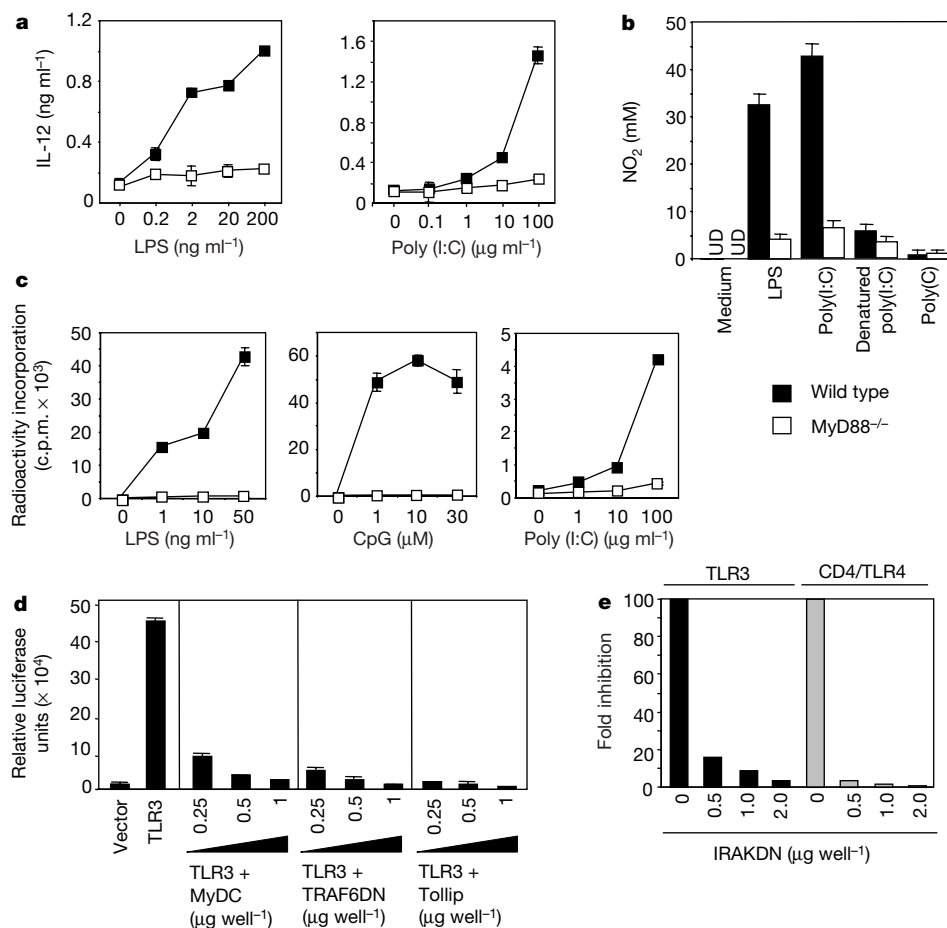


Figure 5 Responses to poly(I:C) in MyD88^{-/-} and TLR3-transfected 293T cells. **a**, Bone-marrow-derived dendritic cells from wild-type or MyD88^{-/-} mice were treated with increasing concentrations of LPS and poly(I:C). After 24 h, IL-12 in the culture supernatants was measured by ELISA. **b**, Bone-marrow-derived macrophages were stimulated with 20 ng ml⁻¹ LPS, 20 μg ml⁻¹ poly(I:C), 20 μg ml⁻¹ denatured poly(I:C) or 20 μg ml⁻¹ poly(C) for 24 h, and the concentration of NO₂ in the culture supernatants was measured. **c**, Total splenocyte suspensions (2 × 10⁶ ml⁻¹) from wild-type or MyD88^{-/-} mice were treated with the indicated amounts of LPS, CpG or poly(I:C) in RPMI, 10% FBS. After 48 h, cells were pulsed with [³H]thymidine for 16 h and radioactivity incorporation

was measured with a β-scintillation counter. **d**, 293T cells were transiently transfected as in Fig. 1a, including the indicated amounts of dominant negative (DN) constructs: MyDC, DN MyD88 with death-domain deletion; TRAF6DN, TRAF6 with deletion of N-terminal interaction domain; and full-length Tollip. Cells were left untreated or stimulated with 25 μg ml⁻¹ poly(I:C) for 8 h (black bars) and luciferase activity was measured. **e**, 293T cells were transfected with TLR3 or a constitutively active CD4/TLR4 chimera, together with an NF- κ B luciferase reporter and the indicated amounts of DN IRAK (IRAKDN). Cells were stimulated with 25 μg ml⁻¹ poly(I:C) for 8 h, and luciferase levels were measured.

minor increase in proliferation compared with wild-type cells when stimulated with poly(I:C) (Fig. 5c). As expected, neither LPS nor CpG induced proliferation of MyD88^{-/-} splenocytes (Fig. 5c). Thus, MyD88 is involved in dsRNA-induced cellular responses.

We next examined the TLR3-induced signal transduction pathway activated by poly(I:C) treatment. Co-transfection of 293T cells with increasing amounts of dominant negative MyD88 or dominant negative TRAF6 together with a constant amount of TLR3 resulted in a dose-dependent inhibition of the stimulation induced by poly(I:C) (Fig. 5d). Moreover, overexpression of Tollip, a protein that interacts with the intracellular domain of the IL-1 receptor and IRAK¹⁸, was also found to profoundly inhibit the poly(I:C)-induced activation of TLR3 (Fig. 5d). Finally, poly(I:C)-induced activation of TLR3 was inhibited by a double negative IRAK in a dose-dependent manner, similarly to the inhibition of signalling by CD4/TLR4, a constitutively active chimera of TLR4 (ref. 19), which was used as a positive control (Fig. 5e). These results suggest that MyD88, IRAK, TRAF6 and Tollip are components of the TLR3-induced signalling pathway.

Previous findings reported by several groups have shown that poly(I:C) treatment of macrophages induces activation of NF- κ B and MAP kinases^{12,20}. We next analysed the degradation of I κ B α and activation of MAP kinase in MyD88^{-/-} cells in response to poly(I:C). The kinetics of I κ B α degradation in MyD88^{-/-} macrophages activated with poly I:C was similar to that in wild-type macrophages (Fig. 6a). By contrast, I κ B α degradation was delayed in MyD88^{-/-} cells treated with LPS or absent in cells treated with CpG (Fig. 6a), as previously reported^{21,22}. Our analysis of stress-activated protein

kinase (SAPK)/JNK and p38 MAP kinases demonstrated that poly(I:C) induced activation of JNK and p38 to a similar extent and with similar kinetics in wild-type and MyD88^{-/-} cells (Fig. 6b). Phosphorylation of JNK and p38 MAP kinase in LPS-treated MyD88^{-/-} macrophages was delayed, whereas in CpG-treated cells it was completely absent (Fig. 6b), consistent with previously published reports^{17,21,22}. Thus, activation of NF- κ B and MAP kinases in response to poly(I:C) can occur in the absence of MyD88.

We next tested the ability of MyD88^{-/-} dendritic cells to mature in response to poly(I:C). Surprisingly, we found that poly(I:C)-induced maturation of dendritic cells, as measured by upregulation of MHC class II and CD86 molecules, was not impaired in MyD88^{-/-} cells (Fig. 6c). Our results suggest that, in response to poly(I:C), TLR3 is involved in cytokine production via a MyD88-dependent pathway, whereas maturation of dendritic cells and activation of NF- κ B and MAP kinases are induced via a MyD88-independent pathway. This dichotomy in signalling responses is similar to that observed for LPS signalling through TLR4 (ref. 23). However, it is not yet clear whether the MyD88-independent signalling induced by TLR4 and TLR3 relies on the same pathway.

We have shown here that TLR3 is critical in the recognition of dsRNA. The presence of viral dsRNA seems to trigger many of the cellular responses to viral infection, through activation of dsRNA-dependent enzymes including PKR, which inhibits viral protein synthesis, and the IFN-inducible 2'-5'-adenylate synthase/RNase L system, which degrades viral RNA^{3,24}. Both of these molecules are intracellular and can initiate the innate immune response when viral infection occurs. However, the fact that mice defective in PKR, and

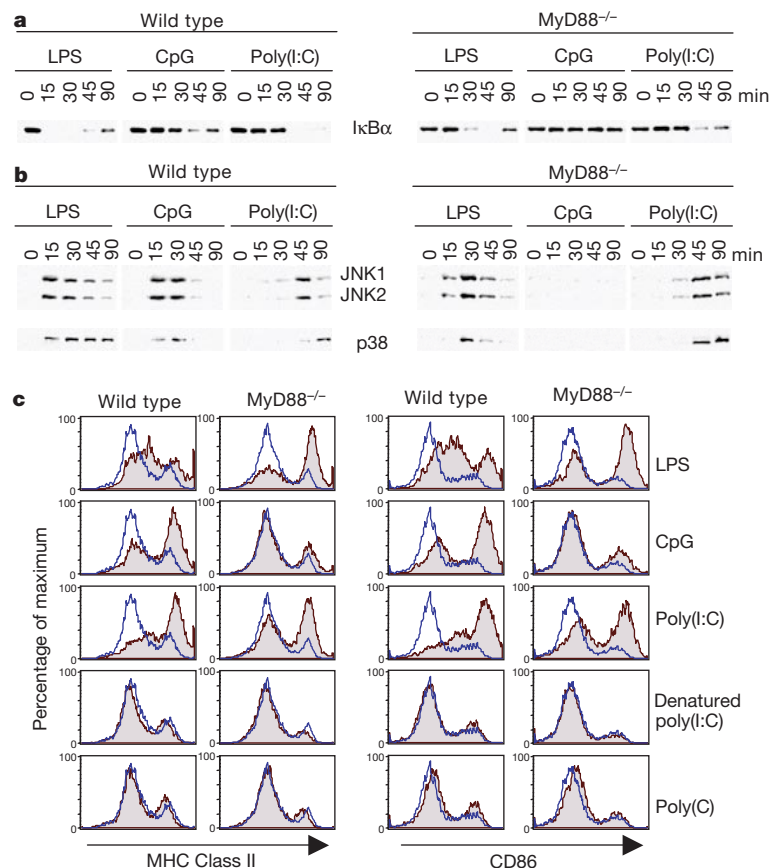


Figure 6 Cellular responses independent of MyD88. **a**, Macrophages from wild-type or MyD88^{-/-} mice were stimulated with 20 ng ml⁻¹ LPS, 10 μ M CpG oligonucleotide or 20 μ g ml⁻¹ poly(I:C) for the indicated durations. Cell lysates were prepared and western blotting was performed with anti-I κ B α antibody. **b**, The same lysates were blotted with phospho-specific antibodies for SAPK/JNK and p38 MAP kinases. **c**, Dendritic cells were

stimulated with 10 ng ml⁻¹ LPS, 10 μ M CpG, 20 μ g ml⁻¹ poly(I:C), poly(C) or denatured poly(I:C) for 24 h and CD11c⁺ cells analysed for the cell surface expression of the indicated molecules by flow cytometry. White area, untreated dendritic cells; grey area, treated dendritic cells.

in both PKR and Rnase L, can still respond to dsRNA or viral infection, suggests that there are additional pathways of viral recognition^{25–27}. Lysis of virally infected cells leads to the release of dsRNA, which then can be detected by the adjacent cells through a transmembrane receptor. Alternatively, after limited proteolysis, dsRNA-containing virions could be recognized in TLR3⁺ vesicles. The importance of TLR3 in the antiviral response, however, remains to be established. Challenge of TLR3-deficient mice with a range of viruses will be required to elucidate whether TLR3 has a role in the host's defence against viruses. □

Methods

Generation of TLR3^{-/-} mice

For the generation of TLR3^{-/-} mice the genomic DNA of the *Tlr3* gene was isolated from a 129SV mouse genomic library (Stratagene). A targeting vector was constructed from a 3.0-kilobase (kb) *NcoI*–*EcoRI* fragment containing the 5' end of the *Tlr3* gene and a 6.7-kb *NcoI*–*XhoI* fragment containing exons 2–4 and the 3' end of the *Tlr3* gene. The first exon of the *Tlr3* gene, including the translation-initiation codon ATG and the signal peptide, was replaced by a *neo* cassette flanked by two *loxP* sites, and a thymidine kinase gene was used for negative selection of clones with random integration of the targeting vector (Fig. 2c). The targeting construct was transfected into embryonic stem cells (W9.5) using a standard protocol, and homologous recombinants were identified by Southern blot analysis (Fig. 2d). Thirteen *Tlr3*-targeted embryonic stem clones were identified and three of them used to generate chimaeric mice. Homozygous TLR3^{-/-} and wild-type (TLR3^{+/+}) mice were generated by intercrossing F₁ heterozygous (TLR3^{+/-}) mice. The homozygous mice were then interbred for the current studies.

Reagents

Poly(I:C) and PGN from *Staphylococcus aureus* were purchased from Amersham and Fluka, respectively. Poly(A:U), poly(C), poly(dI:dC) and LTA from *S. aureus* and LPS from *Salmonella enteritidis* (used in Figs 2 and 3) or *Escherichia coli* (used in Figs 5 and 6) were purchased from Sigma. CpG oligonucleotides were from Keck, Yale University. Viral dsRNA from type I Lang mammalian reovirus was kindly provided by M. Nibert.

NF-κB assays

HEK 293T cells were transiently transfected using Lipofectamine 2000 reagent (Gibco BRL), according to the manufacturer's instruction, with the indicated amounts of expression plasmids and reporter pBIIXLuc plasmid. The total amount of transfected DNA was kept constant by adding empty vector (Fig. 1), or 293T cells were co-transfected with 5 ng Renilla luciferase (pRL-null, Promega) (Fig. 2f). Where indicated, cells were treated with poly(I:C), poly(A:U), poly(C), poly(dI:dC) or PGN. Cells were lysed 6–8 h after stimulation and luciferase activity was measured following the manufacturer's instructions. The Flag-tagged human TLR3 and TLR2 were cloned into pCMV1 Flag vectors. Dominant negative TLR3 and TLR2 have a deletion of the TIR domain. Dominant negative MyD88, IRAK, TRAF6 and CD4/TLR4 were as described previously²⁸. The Tollip expression construct was created by inserting a cDNA fragment, amplified by polymerase chain reaction (PCR), into the pcDNA3 expression vector (Promega).

Analysis of macrophages and dendritic cells

Bone marrow cells were cultured in DMEM medium supplemented with 20% fetal bovine serum (FBS) and 30% supernatant derived from L929 confluent cells. At day 5 or 6, immature macrophages were collected and cultured in the presence or absence of stimuli in RPMI 1640 medium, 5% FBS. Bone marrow dendritic cells were prepared as described previously²⁹. On day 5, adherent cells were stimulated or left untreated and on day 7 they were collected for analysis.

Measurement of cytokine production

Macrophages (1 × 10⁶ cells ml⁻¹), dendritic cells (1 × 10⁶ cells ml⁻¹) or total splenocytes (6 × 10⁶ cells ml⁻¹) were cultured with the indicated stimuli for 24 h. Concentrations of IL-6, IL-12 p40/p70 (antibodies from PharMingen) and TNF-α (Duoset) in the cultured supernatants were measured by enzyme-linked immunosorbent assay (ELISA). NO₂ production was measured by the Griess assay.

Electrophoretic mobility shift assay (EMSA)

Macrophages from wild-type or TLR3^{-/-} mice were stimulated for the indicated periods and then nuclear proteins were extracted. The extracts (3 g) were incubated for 30 min at room temperature with a specific ³²P-labelled probe containing NF-κB DNA-binding sites, electrophoresed on a 5% polyacrylamide gel and visualized by autoradiography.

PCR with reverse transcription (RT-PCR)

Total RNA from macrophages was isolated with TRIzol reagent (Gibco BRL), and contaminant DNA was removed by Dnase (Ambion), according to the manufacturer's instructions. Total RNA (5 μg) was reverse transcribed using Superscript reverse transcriptase (Gibco BRL) in a total volume of 20 μl. Five per cent of this reaction was used as a template for PCR amplification with Tsg DNA polymerase (LAMBDA Biotech) for 25–30 cycles at 94 °C for 30 s, 54 °C for 40 s, and 72 °C for 40 min. The IFN-α, IFN-β and HPRT cDNAs were amplified with the following primers: 5'-CTCGTGATGCTGATAG

TGATGAGC-3' and 5'-CCACACTTTGTCTCACACTCACTCC-3' for IFN-α, 5'-TTCTGCTGTGCTTCTCCAC-3' and 5'-GATTCATACAGTCCAGAGTC-3' for IFN-β, and 5'-GTTGGATACAGCCAGACTTTGTTG-3' and 5'-TCGGATCCGGTCCGATGGGAG-3' for HPRT.

Western blotting

Macrophages (5 × 10⁵ cells ml⁻¹) were activated with the indicated stimuli and for the indicated time and lysed; total protein (80 μg) was resolved on 10% SDS-PAGE gels and transferred to Immobilon P membranes. Blotting was performed with anti-IκBα antibody (Santa Cruz Biotechnology), phospho-SAPK/JNK or phospho-p38 MAPK (New England Biolabs). Bands were visualized with secondary HRP-conjugated antibodies and the ECL System (Amersham Pharmacia).

Received 18 July; accepted 30 August 2001.

1. Akira, S. Toll-like receptors and innate immunity. *Adv. Immunol.* **78**, 1–56 (2001).
2. Medzhitov, R. M. & Janeway, C. A. Innate immune recognition: Mechanisms and pathways. *Immunol. Rev.* **173**, 89–97 (2000).
3. Jacobs, B. L. & Langland, J. O. When two strands are better than one: the mediators and modulators of the cellular responses to double-stranded RNA. *Virology* **219**, 339–349 (1996).
4. Guidotti, L. G. & Chisari, F. V. Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu. Rev. Immunol.* **19**, 65–91 (2001).
5. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–264 (1998).
6. Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
7. Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).
8. Hoshino, K. *et al.* Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J. Immunol.* **162**, 3749–3752 (1999).
9. Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of Gram-negative and Gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
10. Takeuchi, O. *et al.* TLR6: A novel member of an expanding toll-like receptor family. *Gene* **231**, 59–65 (1999).
11. Clemens, M. J. & Elia, A. The double-stranded RNA-dependent protein kinase PKR: structure and function. *J. Interferon Cytokine Res.* **17**, 503–524 (1997).
12. Chu, W. M. *et al.* JNK2 and IKKβ are required for activating the innate response to viral infection. *Immunity* **11**, 721–731 (1999).
13. Maggi, L. B. Jr *et al.* Potential role of PKR in double-stranded RNA-induced macrophage activation. *EMBO J.* **19**, 3630–3638 (2000).
14. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* **282**, 2085–2088 (1998).
15. Kimura, M. *et al.* Comparison of acute phase responses induced in rabbits by lipopolysaccharide and double-stranded RNA. *Am. J. Physiol.* **267**, R1596–R1605 (1994).
16. Kopp, E. B. & Medzhitov, R. The Toll-receptor family and control of innate immunity. *Curr. Opin. Immunol.* **11**, 13–18 (1999).
17. Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
18. Burns, K. *et al.* Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nature Cell Biol.* **2**, 346–351 (2000).
19. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
20. Goh, K. C., deVeer, M. J. & Williams, B. R. The protein kinase PKR is required for p38 MAPK activation and the innate immune response to bacterial endotoxin. *EMBO J.* **19**, 4292–4297 (2000).
21. Hacker, H. *et al.* Immune cell activation by bacterial CpG-DNA through myeloid differentiation marker 88 and tumor necrosis factor receptor-associated factor (TRAF)6. *J. Exp. Med.* **192**, 595–600 (2000).
22. Schnare, M., Holt, A. C., Takeda, K., Akira, S. & Medzhitov, R. Recognition of CpG DNA is mediated by signaling pathways dependent on the adaptor protein MyD88. *Curr. Biol.* **10**, 1139–1142 (2000).
23. Akira, S. Toll-like receptors: lessons from knockout mice. *Biochem. Soc. Trans.* **28**, 551–556 (2000).
24. Kumar, M. & Carmichael, G. G. Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol. Mol. Biol. Rev.* **62**, 1415–1434 (1998).
25. Abraham, N. *et al.* Characterization of transgenic mice with targeted disruption of the catalytic domain of the double-stranded RNA-dependent protein kinase, PKR. *J. Biol. Chem.* **274**, 5953–5962 (1999).
26. Yang, Y. L. *et al.* Deficient signaling in mice devoid of double-stranded RNA-dependent protein kinase. *EMBO J.* **14**, 6095–6106 (1995).
27. Zhou, A., Paranjape, J. M., Der, S. D., Williams, B. R. & Silverman, R. H. Interferon action in triply deficient mice reveals the existence of alternative antiviral pathways. *Virology* **258**, 435–440 (1999).
28. Medzhitov, R. *et al.* MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol. Cell* **2**, 253–258 (1998).
29. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176**, 1693–1702 (1992).

Acknowledgements

We thank M. Nibert for providing viral dsRNA; S. Akira for providing the TLR2^{-/-} and MyD88^{-/-} mice; L. Evangelisti, D. Butkus, C. Hughes and J. Stein for technical assistance; and F. Manzo for secretarial work. This work was supported by the Howard Hughes Medical Institute and grants from the National Institutes of Health (NIH). L.A. was the recipient of a Human Frontier Science Program Postdoctoral Fellowship. R.M. is a Searle Scholar, R.A.F. is an Investigator and R.M. is an Assistant Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for material should be addressed to R.A.F. (e-mail: richard.flavell@yale.edu) or R.M. (e-mail: ruslan@yale.edu). The murine *TLR3* sequence has been deposited in GenBank under accession number AF420279.

Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in *Drosophila*

Joan S. Steffan^{*}, Laszlo Bodai[†], Judit Pallos[‡], Marnix Poelman^{*}, Alexander McCampbell[§], Barbara L. Apostol^{*}, Alexsey Kazantsev[§], Emily Schmidt^{*}, Ya-Zhen Zhu^{*}, Marilee Greenwald^{*}, Riki Kurokawa^{||}, David E. Housman[§], George R. Jackson[¶], J. Lawrence Marsh[†] & Leslie M. Thompson^{*}

^{*} Department of Psychiatry and Human Behavior, Gillespie 2121, University of California, Irvine, California 92697, USA

[†] Department of Developmental and Cell Biology, University of California, Irvine, California 92697, USA

[‡] Neurogenetics Branch, NINDS, NIH, 10/3B14, 10 Center Drive, MSC 1250, Bethesda, Maryland 20892, USA

[§] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

^{||} Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA

[¶] Department of Neurology, University of California, Los Angeles, California 90095, USA

Proteins with expanded polyglutamine repeats cause Huntington's disease and other neurodegenerative diseases. Transcriptional dysregulation and loss of function of transcriptional co-activator proteins have been implicated in the pathogenesis of these diseases¹. Huntington's disease is caused by expansion of a repeated sequence of the amino acid glutamine in the abnormal protein huntingtin (Htt). Here we show that the polyglutamine-

containing domain of Htt, Htt exon 1 protein (Httex1p), directly binds the acetyltransferase domains of two distinct proteins: CREB-binding protein (CBP) and p300/CBP-associated factor (P/CAF). In cell-free assays, Httex1p also inhibits the acetyltransferase activity of at least three enzymes: p300, P/CAF and CBP. Expression of Httex1p in cultured cells reduces the level of the acetylated histones H3 and H4, and this reduction can be reversed by administering inhibitors of histone deacetylase (HDAC). *In vivo*, HDAC inhibitors arrest ongoing progressive neuronal degeneration induced by polyglutamine repeat expansion, and they reduce lethality in two *Drosophila* models of polyglutamine disease. These findings raise the possibility that therapy with HDAC inhibitors may slow or prevent the progressive neurodegeneration seen in Huntington's disease and other polyglutamine-repeat diseases, even after the onset of symptoms.

Huntington's disease is a late-onset neurodegenerative disease characterized by a movement disorder, neuropsychiatric symptoms and cognitive deficits caused by expansion of a glutamine repeat in the Htt protein. Currently, no cure or effective treatment for this agonizing, lethal disease exists. The aetiology of several other neurodegenerative diseases, including spinocerebellar ataxia 1 and Kennedy's disease, also involves expansion of a polyglutamine repeat². The pathology of these diseases may involve transcriptional dysregulation¹. We previously found that a fragment of mutant Htt interacts directly with CBP, which contains an acetyltransferase domain and is a co-activator of numerous promoters³⁻⁵. Mutant Htt represses transcription from CBP/p300 co-activated promoters in cell culture^{5,6}, and CBP and other transcriptional regulatory proteins are sequestered in cytoplasmic and nuclear aggregates in both transgenic mice and patient brains^{5,6}. In addition, ectopic overexpression of CBP reduces polyglutamine-mediated death of cultured cells^{6,7}. These observations prompted us to investigate whether Htt interacts with other acetyltransferase-containing enzymes, to

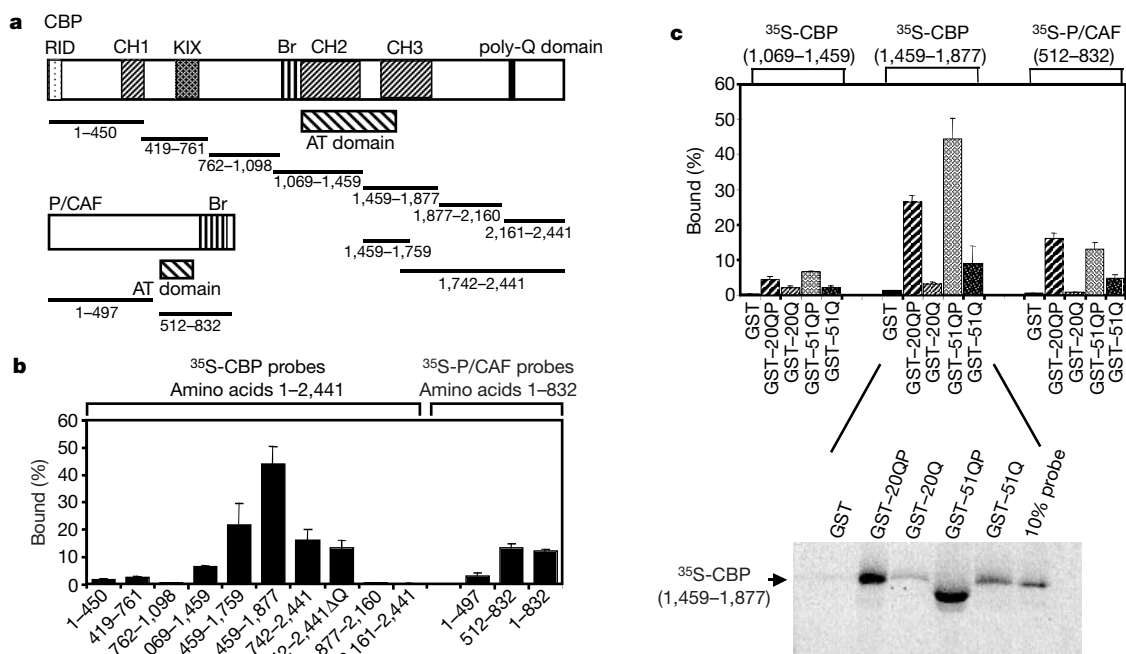


Figure 1 Htt interacts with the acetyltransferase (AT) domain of CBP and P/CAF *in vitro*. **a**, Schematic representations of CBP and P/CAF. RID, nuclear hormone receptor interacting domain; CH1, CH2, CH3, cysteine-histidine-rich regions 1, 2 and 3; Br, bromodomain; KIX, CREB-binding domain. The amino-acid residues used as ³⁵S-labelled probes for GST pull-down assays with GST-Htt proteins are designated below each protein diagram. **b**, Results of GST pull-down assays using radioactive domains of CBP and P/CAF with GST-51QP Htt. The AT domains of both CBP and P/CAF interact with

GST-51QP. **c**, Interaction of GST-Htt fusion proteins, containing wild-type and expanded polyglutamine stretches, with and without the proline-rich domain, with radioactive probes containing the AT domains of CBP and P/CAF. A representative autoradiogram of the middle panel is shown. Slight alteration in the pattern of CBP(1,459-1,877) in the GST-51QP lane as compared with the other lanes is due to co-migration of GST-51QP with labelled CBP(1,459-1,877).

investigate the biochemical basis of the interaction and, finally, to ask whether effects on the levels of acetylation activity are important in polyglutamine-induced neurodegeneration.

The exact nature of the pathogenic entity of Htt is not known, although evidence suggests that cleavage products of mutant Htt containing the expanded polyglutamine domain are the pathogenic agents^{8,9}. Httex1p, a polypeptide encoded by the first exon of Htt that contains the polyglutamines plus 65 additional amino acids, has been shown to cause pathology in mice similar to that of Huntington's disease¹⁰. That this truncated Htt with an expanded polyglutamine domain interacts with full-length CBP *in vitro*⁵ and co-aggregates in cell culture^{5,6,11} suggests two hypotheses for Htt action that are not mutually exclusive. First, it is possible that the available activity of cellular proteins containing natural polyglutamine domains could be reduced by trapping them in aggregates through polyglutamine–polyglutamine interactions. Alternatively, soluble interactions between Htt and cellular targets could directly affect enzymatic activities.

To distinguish the contributions of polyglutamine–polyglutamine interactions from those of other interactions, we investigated which functional domains of CBP are targeted for binding by Httex1p, using glutathione S-transferase (GST) pull-down assays. ³⁵S-labelled protein probes spanning the length of CBP (Fig. 1a) were incubated with glutathione-agarose beads coupled to Httex1p with 51 glutamine repeats and the proline-rich domain. Significant binding was observed to probes containing the CH2 domain and the amino-terminal portion of the acetyltransferase domain (residues 1,069–1,459), to probes containing the carboxy-terminal portion of the acetyltransferase domain (residues 1,459–1,759), and to probes containing part of the acetyltransferase domain as well as the CH3 domain (residues 1,459–1,877) (Fig. 1b). A CBP probe containing part of the CH3 domain and the C-terminal polyglutamine stretch (residues 1,742–2,441) binds to GST–51QP, whereas deletion of the polyglutamine stretch (residues 1,742–2,441ΔQ) has a negligible effect on binding. A CBP fragment containing the C-terminal polyglutamine domain of CBP (residues 2,161–2,441) did not interact *in vitro* with GST–51QP. Our data indicate that the direct interaction between the 51QP Htt fragment and CBP is not mediated by glutamine–glutamine interactions. Rather, they indicate that 51QP interacts primarily with the region of CBP containing the acetyltransferase and CH3 domains.

We next investigated whether Httex1p binds other proteins containing acetyltransferase domains, such as P/CAF. Unlike CBP, P/CAF does not contain a polyglutamine-rich region (Fig. 1a). Nevertheless, full-length P/CAF (residues 1–832), as well as residues containing the acetyltransferase domain (residues 512–832), showed significant (12–13%) binding to GST–51QP *in vitro* (Fig. 1b).

Having found that the expanded repeat domain of Htt interacts with both CBP and P/CAF primarily through the acetyltransferase and neighbouring domains, we sought to determine the features of Httex1p that are important for binding. We tested whether the length of the polyglutamine tract or the presence of the proline-rich region affected this interaction. Binding of GST–Httex1p to the acetyltransferase/CH3 domain of CBP is enhanced when the polyglutamine repeat expands from the normal range (GST–25QP or GST–25Q) to a pathogenic range (GST–51QP or GST–51Q) (Fig. 1c). GST fusions that contain the Htt proline-rich domain (GST–20QP and GST–51QP) interact better with the acetyltransferase-domain probes than do their counterparts without the proline-rich domain (GST–20Q and GST–51Q). Thus, the proline-rich region enhances, but is not essential for, the *in vitro* interaction. A representative autoradiogram and quantification by phosphorimager analysis show binding of CBP probe 1,459–1,877 to the different GST–Htt constructs (Fig. 1c).

Because Httex1p interacts *in vitro* with the acetyltransferase domains of P/CAF and CBP, we measured the effect of Httex1p

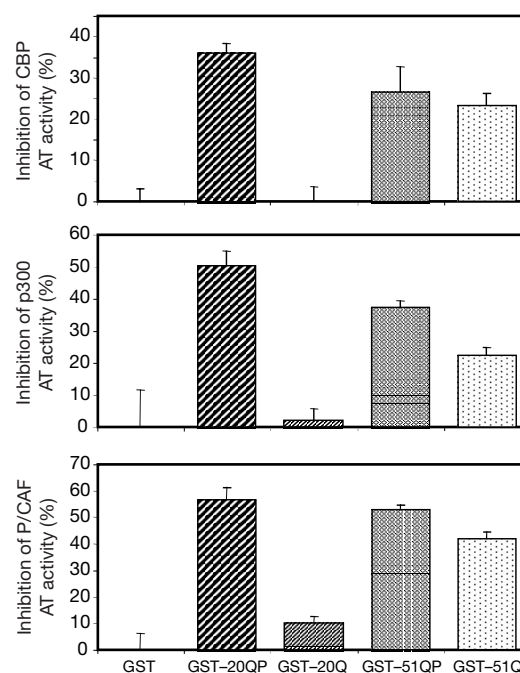


Figure 2 Htt inhibits the acetyltransferase activity of CBP, p300 and P/CAF *in vitro*. GST fusion proteins GST–20QP, GST–51QP and GST–51Q inhibit the acetyltransferase activity of GST–CBP (amino acids 1,099–1,877), GST–p300 (amino acids 1,195–1,707) and GST–P/CAF (amino acids 87–832) on biotinylated H4 peptide.

on the acetyltransferase activity of several widely expressed transcriptional co-activator proteins *in vitro*: CBP, p300 and P/CAF. Equimolar amounts of GST, GST–20QP, GST–20Q, GST–51QP and GST–51Q proteins were incubated with GST–CBP–AT (acetyltransferase) (residues 1,099–1,877), GST–p300–AT (residues 1,195–1,707), or GST–P/CAF–AT (residues 87–832), containing the acetyltransferase domains. Following a short incubation, the acetylation of an H4 peptide was monitored. The presence of the Htt GST fusion protein, GST–51Q, as well as GST–20QP and GST–51QP, which contained the proline-rich region, significantly reduced acetyltransferase activity of CBP, p300 and P/CAF, whereas GST and GST–20Q did not (Fig. 2). We conclude that the direct interaction of Htt with acetyltransferase domains (Fig. 1c) inhibits acetyltransferase function and that long polyglutamine stretches are more potent than short normal-range repeats. The presence of the proline-rich region increases the inhibitory potential of polyglutamine polypeptides. Surprisingly, GST–51Q inhibits acetyltransferase activity almost as well as GST–51QP, demonstrating that expanded polyglutamine stretches can interact with acetyltransferase domains and inhibit function without the proline-rich region.

Because Httex1p interacts directly with acetyltransferase domains and inhibits their function *in vitro*, we analysed acetylation levels of H3 and H4 in two PC12 cell lines stably transfected with inducible Httex1p constructs (our own unpublished data and E. Schweitzer, unpublished data). Expression of Httex1p with a normal-range repeat (25QP) reduces acetylation of both H3 and H4 compared with uninduced cells (Fig. 3a). Expression of an expanded repeat Httex1p (103QP) shows an even greater reduction of acetylation (Fig. 3a). To verify these results, we used a second transformed cell line, generated in a different strain of PC12 cells. Induction of an expanded polyglutamine repeat Httex1p (103QP) in the second PC12 line also leads to reduction of H4 acetylation in whole-cell extracts compared with uninduced cells (Fig. 3b). When these cells are treated with the HDAC inhibitor sodium butyrate, levels of H4 acetylation in both uninduced and induced cells are increased (Fig. 3b). Trichostatin A (TSA) and suberoylanilide hydroxamic

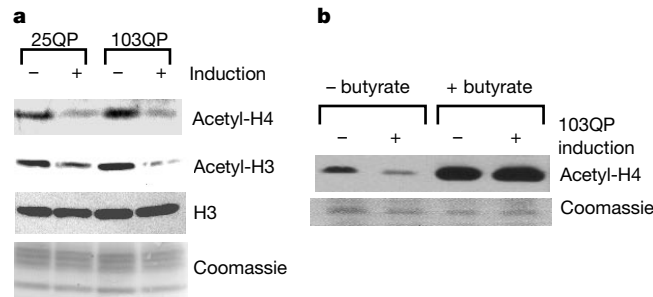


Figure 3 A reduction in acetylation of histones H3 and H4 is induced by Htt expression in cell culture. **a**, PC12 cells induced to express 25QP-GFP or 103QP-GFP show a reduction in the level of acetylation of both histones H3 and H4. **b**, A second strain of stably transfected PC12 cells, grown in the absence of butyrate but induced to express F103QP-

EGFP, show a reduction in the level of acetylation of H4, which can be reversed by butyrate treatment. Equivalent levels of histones are shown by Coomassie blue staining (**a** and **b**) or with anti-histone H3 (**a**).

acid (SAHA) also increased acetylation of H3 and H4, reversing the decrease induced by Httex1p (data not shown). Although acetylation levels are altered, total histone levels in whole-cell extracts, determined by Coomassie blue staining of total cell lysates, are unchanged by induction of Htt proteins. Thus, expression of Httex1p causes a global reduction in acetylation of histones H3 and H4 that is reversed in the presence of HDAC inhibitors.

We found that even normal repeat Htt can bind to and inhibit acetyltransferase activity (Figs 1 and 2), so why is pathology associated with only expanded repeat Htt? Mutant Htt can be proteolytically processed; the pathogenic fragment localizes to the nucleus, where it is capable of inhibiting acetyltransferase activity^{8,12,13}. However, unexpanded repeat Htt does not normally localize to the nucleus¹⁴ and therefore is not present in the cellular compartment appropriate to inhibit nuclear CBP activity. Indeed, unexpanded polyglutamine repeats can cause pathology. For instance, unexpanded human ataxin-1 protein, which contains 30 glutamines, is normally localized to the nucleus¹⁵. If it is expressed at sufficiently high levels there, it can produce neurodegenerative phenotypes similar to expanded 82-glutamine ataxin-1 in either *Drosophila* or mice¹⁵. Thus, polyglutamine pathogenesis depends heavily on both level and location.

The reduced acetylation of histones observed in the presence of expanded repeat Httex1p *in vitro* (with or without prolines), and the subsequent reversal of this effect with HDAC inhibitors in cell culture, suggested that reduced acetyltransferase activity may be an important component of polyglutamine pathogenesis *in vivo*. Expanded polyglutamine peptides alone¹⁶ as well as expanded repeat Httex1p (this report) are intrinsically cytotoxic and cause reduced viability and neuronal degeneration when expressed in *Drosophila* neurons. If polyglutamine pathology involves suppression of histone acetylation, then one would predict that inhibition of the deacetylation process by either of two completely independent mechanisms (for example, pharmacologically or genetically) would slow or reduce polyglutamine pathogenesis *in vivo*. To test this hypothesis, transgenic flies were engineered to express either Httex1p or just polyglutamine peptides¹⁶ in neurons. We monitored the effects of feeding flies with the HDAC inhibitors butyrate and SAHA, and of genetically reducing the activity of their HDAC, on both lethality and degeneration of photoreceptor neurons.

Neurodegeneration is most readily observed in the fly compound eye, which is composed of a regular trapezoidal arrangement of seven visible rhabdomeres (subcellular light-gathering structures) produced by the photoreceptor neurons of each *Drosophila* ommatidium (Fig. 4j). We found that expression of Httex1p with 93 glutamines (Q93) led to a progressive loss of rhabdomeres (Fig. 4a). Rather than the normal seven visible rhabdomeres, the number of rhabdomeres seen in flies expressing Httex1p Q93 progressively declined from an average of 6.35 at day 1 to 5.13 and 4.66 at days 6

and 12 after eclosion, respectively (that is, following emergence from the pupal case as an adult). Rearing larvae that expressed Httex1p Q93 on SAHA- or butyrate-containing food reduced the level of degeneration observed (Fig. 4b, c). Expression of the Httex1p Q93 transgene results in ~70% lethality (data not shown) and early adult death (Fig. 4d). In contrast, animals fed the HDAC inhibitor SAHA show increased viability (10 μ M SAHA suppresses lethality to 45%; data not shown), and early adult death is markedly repressed in a concentration-dependent manner (Fig. 4d).

The effects of HDAC inhibitors on transgenic flies expressing extended polyglutamine peptides alone (Q48) were similar to those described above: Q48 flies fed butyrate or SAHA had the same distribution of rhabdomeres by day 6 as 1-day-old flies (data not shown), whereas their siblings that did not receive HDAC inhibitors showed a significant degeneration of rhabdomere number over time (average of 5.47 at day 1 versus 3.92 at day 6; Fig. 4g, j). Even when fed HDAC inhibitors only after emerging from the pupal case as adults, progressive degeneration of photoreceptor neurons was still prevented (Fig. 4h, i). Therefore, even when administered to animals already exhibiting neurodegeneration, HDAC inhibitors markedly retard (or arrest) further neuronal degeneration. Thus, HDAC inhibitors rescue pathological effects of both polyglutamine peptides and Htt exon 1 polypeptides *in vivo*.

It is possible that HDAC inhibitors might affect cellular processes other than the deacetylase pathways. As an independent test of the significance of acetylation levels in the pathogenic process, we manipulated acetylation levels genetically and examined pathology. The *Drosophila* Sin3A locus encodes a co-repressor protein that is a component of HDAC complexes¹⁷. We found that reducing the levels of HDAC by a partial loss of function mutant, Sin3A⁰⁸²⁶⁹, in heterozygotes increased the viability of Httex1p Q93 flies from 29% to 65% and led to a reduction in the extent and rate of neurodegeneration (Fig. 4e). Because this mutant allele represents a partial loss of Sin3A function, the effect on rescue of neurodegeneration may be less than that observed in the presence of HDAC inhibitors. Thus, both genetic and pharmacological reductions in the activity of HDAC reduce the rate and extent of polyglutamine-induced pathology.

To rule out the possibility that the rescue of degeneration and lethality by HDAC inhibitors was simply due to altered expression of the polyglutamine transgenes in the presence of HDAC inhibitors, we used western blotting to compare the level of transgene expression in larvae expressing exon 1 polypeptides in neurons in the presence and absence of butyrate or SAHA (Fig. 4f). Transgene expression was unaltered by the presence of HDAC inhibitors.

Although we have shown functional interactions of Htt with CBP, P/CAF and p300, these results do not exclude the possibility that other acetyltransferases may be targeted, but they do suggest that

treatments that raise global levels of acetylation may be effective in ameliorating the effects of Huntington's disease and other neurodegenerative processes, even after the onset of symptoms.

The above results raise the possibility that Htt peptides can lead to reduced levels of acetylation and transcription, both by binding to acetyltransferase domains and inhibiting soluble activity, and by sequestering polyglutamine-containing transcription factors (such

as CBP and others) by trapping them into aggregates. When tested *in vivo* in *Drosophila* models of polyglutamine pathogenesis, inhibition of the deacetylation process by two independent mechanisms—pharmacological (HDAC inhibitors) and genetic (reduction of Sin3A activity)—reduced degeneration of photoreceptor neurons and lethality. Other investigators have observed a reduction in acetylation induced by expanded polyglutamine-repeat proteins,

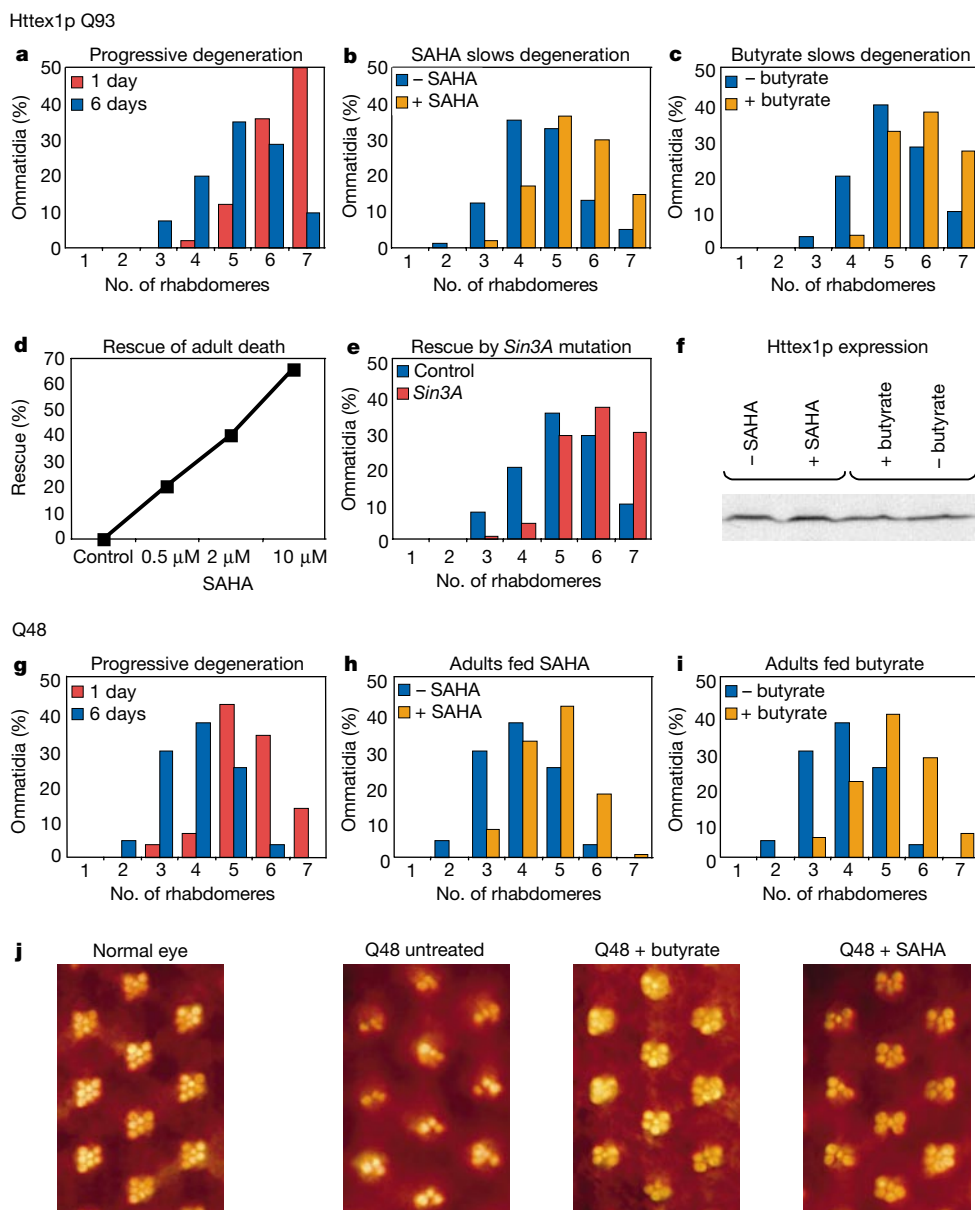


Figure 4 *In vivo* consequences of altered acetyltransferase function. **a**, The number of rhabdomeres per ommatidium at 1 and 6 d after eclosion in flies expressing Httex1p Q93 shows progressive loss. **b**, Administration of SAHA slows photoreceptor degeneration. The number of rhabdomeres per ommatidium at 6 d after eclosion is markedly improved in Httex1p-Q93-expressing flies fed SAHA. Animals were fed 0.5, 2 and 10 μ M SAHA. Results with 2 μ M SAHA are shown. **c**, Administration of butyrate slows neuronal degeneration. The number of rhabdomeres per ommatidium at 6 d after eclosion is markedly improved in Httex1p-Q93-expressing flies fed butyrate. Animals were fed 10, 30 and 100 mM butyrate. Results with 100 mM butyrate are shown. **d**, Administration of SAHA improves the 6-d survival of adult flies expressing Httex1p. Animals were fed 0.5, 2 and 10 μ M SAHA dissolved in DMSO. Per cent rescue was calculated as follows: (per cent surviving – per cent surviving on solvent alone)/(1 – per cent surviving on solvent alone). At least 100 flies were evaluated for lethality per genotype. **e**, Genetically reducing

deacetylase activity in *Sin3A* heterozygotes slows degeneration. Flies expressing Httex1p Q93 and heterozygous for a *Sin3A* mutation were compared with similar flies without the *Sin3A* mutation. The photoreceptor distribution was monitored at 6 d. **f**, Expression of the Httex1p Q93 transgene is unchanged by treatment with either SAHA or butyrate. A western blot of extracts from larvae expressing Httex1p Q93 and treated either with solvent alone or solvent with SAHA or butyrate was probed with anti-Htt antibody. Similar amounts of protein were loaded, as determined by Bradford assays and confirmed by Coomassie staining of the gel. **g**, Rhabdomeres in the eyes of flies expressing tagged Q48 peptides also show progressive loss. **h**, **i**, Progressive degeneration of photoreceptor neurons in Q48-expressing flies is arrested by 2 μ M SAHA (**h**) or 100 mM butyrate (**i**) even when feeding is initiated only in adult flies. **j**, Photographs of ommatidia from Q48-expressing flies with and without HDAC inhibitors.

for example, in yeast (R. Hughes and S. Fields, personal communication) and in a cell-culture model of Kennedy's disease (A. McCampbell and K. Fischbeck, personal communication). These results strongly implicate the state of acetylation in the pathogenic process. Because several HDAC inhibitors, including SAHA¹⁸, are currently approved by the US Food and Drug Administration (FDA) for use in other clinical settings or are in phase I clinical trials, HDAC inhibitors should be seriously considered as potential therapeutic agents for Huntington's disease and related diseases. □

Methods

Plasmid constructs

GST-Htt exon 1 fusion proteins¹⁹, pcDNA3.1 containing the complementary DNA for CBP¹¹, pGST-P/CAF-AT domain (encoding amino acids 87–768), and pcDNA3 containing the cDNA for P/CAF²⁰ were used. A Flag tag was cloned in front of alternating CAG/CAA repeats¹¹ to create F103QP-EGFP. For transgenic *Drosophila* lines, Htt exon1 constructs¹⁹ were subcloned into pUAST²¹.

GST pull-down assays

GST fusion proteins were purified and experiments were performed as described⁵. Domains of CBP were amplified by polymerase chain reaction (PCR) and the fragments cloned into pcDNA3.1. Binding and activity data were compared by an analysis of variance with StatView.

Acetyltransferase assays

The effect of Htt proteins on the acetyltransferase activity was assayed *in vitro* by a modified technique²². GST fusion proteins were washed in 1× HAT buffer (10 mM butyrate at pH 7.5, 10% glycerol, 50 mM Tris HCl at pH 8.0, 0.5 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.1 mM phenylmethyl sulphonyl fluoride (PMSF)), then eluted with 15 mM glutathione in 1× HAT buffer containing 50 mM NaCl. Quantified by Coomassie stain, 0.6 nmol of Htt GST fusion proteins or GST alone were incubated for 10 min at room temperature with 10 pmol of GST–CBP (amino acids 1,099–1,877), 4 pmol of GST–p300 (amino acids 1,195–1,707) (purchased from Upstate Biotechnology), or 360 pmol of GST–P/CAF (amino acids 87–832), as estimated by Bradford analysis, in a total volume of 55 µl. We then added 10 nCi [¹⁴C]-acetyl coenzyme A (52 mCi mmol⁻¹, NEN) and 2 µg of biotinylated N-terminal H4 (amino acids 1–21) peptide (Upstate Biotechnology) in a volume of 4 µl, and the mixture was incubated for 45 min at 30 °C. We added 500 µl of 1× HAT buffer with 30 µl of a 50% slurry of streptavidin-agarose (Upstate Biotechnology) pre-equilibrated in 1× HAT buffer. This mixture was rotated at 4 °C for 20 min, then spun in a microfuge at 2,600 g for 2 min. The supernatant was removed, the pellet washed twice in RIPA buffer (50 mM Tris at pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA and 0.1 mM PMSF) and counted in a liquid scintillation counter. Assays were done in quadruplicate and the standard error of the mean calculated.

Histone acetylation analysis of PC12 cells

PC12 cells, stably transfected with ecdysone-inducible (Invitrogen) F103QP-EGFP (ref. 5 and our own unpublished data) were plated, induced with 5 µM ponasterone (Invitrogen) for 48 h, and treated for the last 24 h with 5 mM sodium butyrate. Controls included uninduced and/or non-butyrate-treated cells. Cells were lysed in 1× HAT buffer with 50 mM NaCl, 0.3% NP40 and a protease-inhibitor cocktail on ice for 10 min. Fifty micrograms of whole-cell extract were analysed by western blot. Separate, independently derived, ecdysone-inducible PC12 cell lines were also analysed. These cells, denoted PC12/pBWN:Htt ex1Q103-EGFP cells and PC12/pWN:Httex1Q25-EGFP, were a gift of E. S. Schweitzer. These clonal cells contain a modified exon1 of Htt¹¹ inserted into an expression vector containing the *Bombyx* ecdysone-regulated element²³. Twenty-five micrograms of whole-cell extracts from PC12 cells stably transfected with plasmids encoding inducible 25QP-EGFP and 103QP-EGFP were analysed for acetyltransferase activity. Control cells (uninduced) and cells induced with 1 µM tubufenizide for 12 h were analysed by western blot. Anti-acetylated histone H4, anti-histone H3 and anti-acetylated histone H3 (all from Upstate Biotechnology) were used to determine levels of acetylated histones H3 and H4 relative to levels of total H3.

Drosophila stocks and crosses

Expression of polyglutamine-containing peptides is driven by the bipartite expression system upstream activator sequence (UAS)–GAL4 in transgenic flies¹⁶. Injection of plasmids expressing Httex1p with 20 or 93 glutamines produced nine and ten lines, respectively. Two lines with the most severe neuronal phenotype (P463 and P465) were chosen for the studies reported here. Polyglutamine stocks are w; P(w^{+mC} = UAS-Q93htt exon1)^{4F1} and w; P(w^{+mC} = UAS-Q48 + mycflag)²⁰. Constructs under the control of a yeast UAS were crossed to flies expressing the yeast GAL4 transcriptional activator²¹ driven by the neuron-specific promoter *elav* (chromosome-2 driver for Q48 lines²⁴ and X-chromosome driver for Htt exon 1 lines²⁵) that is expressed in all neurons from embryogenesis onwards: w; P(w^{+mC}; w+; *elav* – GAL4)/CyO, P(w^{+mC} = *Act* – GFP)/JMR1CyO *actin*-GFP or w; P(w^{+mW} = *GawB*)/*elav*C155. The inhibitor concentration ranges tested were based on cell culture experiments (SAHA, a gift from Calbiochem) or published position effect variegation studies (butyrate)²⁶. For testing the

effect of *Sin3A* on polyglutamine phenotypes, w; P(w^{+mC} = UAS-Q93htt exon1)^{4F1} virgins were crossed to w; P(w^{+mC}; w Pw; *elav*-GAL4)/Y; *Sin3A*⁹⁸²⁶⁹/Bc *Gla* males.

Pseudopupil analysis and transgene expression

Pseudopupil analysis allows visualization of the arrangement of rhabdomeres in the ommatidia of the compound eye²⁷. Adult flies were decapitated, one eye was dipped in a drop of nail polish and the head was mounted on a microscope slide. Eyes were analysed with a Nikon EFD-3/Optiphot-2 scope using a ×50 objective, and photographed with a Spot camera. At least 200 ommatidia were scored for each condition. For western analysis of transgene expression, 20 larvae from each sample were ground in a buffer containing 0.1 M phosphate at pH7.1, 0.3 M sucrose, 0.02 mM phenylthiourea, protease cocktail and 0.1 mM PMSF, and 200 µg of total lysate were loaded per lane.

Received 14 May; accepted 30 August 2001.

1. Cha, J. H. Transcriptional dysregulation in Huntington's disease. *Trends Neurosci.* **23**, 387–392 (2000).
2. Zoghbi, H. Y. & Orr, H. T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* **23**, 217–247 (2000).
3. Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**, 1553–1577 (2000).
4. Giordano, A. & Avantaggiati, M. L. p300 and CBP: partners for life and death. *J. Cell. Physiol.* **181**, 218–230 (1999).
5. Steffan, J. S. *et al.* The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc. Natl Acad. Sci. USA* **97**, 6763–6768 (2000).
6. Nucifora, F. C. *et al.* Interference by Huntingtin and Atrophin-1 with CBP-mediated transcription leading to cellular toxicity. *Science* **291**, 2423–2428 (2001).
7. McCampbell, A. *et al.* CREB-binding protein sequestration by expanded polyglutamine. *Hum. Mol. Gen.* **9**, 2197–2202 (2000).
8. Paulson, H. L. Protein fate in neurodegenerative proteinopathies: polyglutamine diseases join the (mis)fold. *Am. J. Hum. Genet.* **64**, 339–345 (1999).
9. Wanker, E. E. Protein aggregation and pathogenesis of Huntington's disease: mechanisms and correlations. *Biol. Chem.* **381**, 937–942 (2000).
10. Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493–506 (1996).
11. Kazantsev, A., Preisinger, E., Dranovsky, A., Goldgaber, D. & Housman, D. Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. *Proc. Natl Acad. Sci. USA* **96**, 11404–11409 (1999).
12. Klement, I. A. *et al.* Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* **95**, 41–53 (1998).
13. Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55–66 (1998).
14. Sapp, E. *et al.* Huntingtin localization in brains of normal and Huntington's disease patients. *Ann. Neurol.* **42**, 604–611 (1997).
15. Fernandez-Funez, P. *et al.* Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature* **408**, 101–106 (2000).
16. Marsh, J. L. *et al.* Expanded polyglutamine peptides alone are intrinsically cytotoxic and cause neurodegeneration in *Drosophila*. *Hum. Mol. Gen.* **9**, 13–25 (2000).
17. Neufeld, T. P., Tang, A. H. & Rubin, G. M. A genetic screen to identify components of the *sina* signaling pathway in *Drosophila* eye development. *Genetics* **148**, 277–286 (1998).
18. Marks, P. A., Richon, V. M. & Rifkind, R. A. Histone deacetylase inhibitors: Inducers of differentiation or apoptosis of transformed cells. *J. Natl Cancer Inst.* **92**, 1210–1216 (2000).
19. Scherzinger, E. *et al.* Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: Implications for Huntington's disease pathology. *Proc. Natl Acad. Sci. USA* **96**, 4604–4609 (1999).
20. Yang, X. J., Ogryzko, V. V., Nishikawa, J., Howard, B. H. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).
21. Brand, A. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
22. Ail-Si-Ali, S., Ramirez, S., Robin, P., Trouche, D. & Harel-Bellan, A. A rapid and sensitive assay for histone acetyl-transferase activity. *Nucleic Acids Res.* **26**, 3869–3870 (1998).
23. Suhr, S. T., Gil, E. B., Senut, M. C. & Gage, F. H. High level transactivation by a modified *Bombyx* ecdysone receptor in mammalian cells without exogenous retinoid X receptor. *Proc. Natl Acad. Sci. USA* **95**, 7999–8004 (1998).
24. Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
25. Lin, D. M. & Goodman, C. S. Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. *Neuron* **13**, 507–523 (1994).
26. Reuter, G., Dorn, R. & Hoffman, H. J. Butyrate sensitive suppressor of position-effect variegation mutants of *Drosophila melanogaster*. *Mol. Gen. Genet.* **188**, 480–485 (1982).
27. Franceschini, N. in *Information Processing in the Visual Systems of Arthropods* (ed. Wehner, R.) 75–82 (Springer, Berlin, 1972).

Acknowledgements

We thank K. Fischbeck for discussions and information regarding SAHA, E. Wanker for the GST–Htt fusion protein constructs, and E. Schweitzer for the gift of PC12 cell pellets. This work was supported by the Hereditary Disease Foundation's Cure HD Initiative, the Huntington's Disease Society of America, Human Frontier Science Program and grants from the National Institutes of Health. This work was made possible in part by access to the Optical Biology Shared Resource of the Cancer Center Support Grant at the University of California, Irvine, and the National *Drosophila* Stock Center.

Correspondence and requests for materials should be addressed to L.M.T. (e-mail: lmthomps@uci.edu).

Forkhead transcription factors contribute to execution of the mitotic programme in mammals

Beatriz Alvarez*, Carlos Martínez-A*, Boudewijn M. T. Burgering† & Ana C. Carrera*

* Department of Immunology and Oncology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Universidad Autónoma de Madrid, Cantoblanco, Madrid E-28049, Spain

† Department of Hematology, University Medical Center, Utrecht 3584 CG, The Netherlands

Cell cycle progression is a process that is tightly controlled by internal and external signals. Environmental cues, such as those provided by growth factors, activate early signals that promote cell cycle entry^{1–3}. Cells that have progressed past the restriction point become independent of growth factors, and cell cycle progression is then controlled endogenously. The phosphatidylinositol 3OH kinase (PI(3)K)/protein kinase B (PKB) pathway must be activated in G1 to inactivate forkhead transcription factors (FKH-TFs)^{4,5} and allow cell cycle entry^{2,3}. Here we show that subsequent attenuation of the PI(3)K/PKB pathway is required to allow transcriptional activation of FKH-TF in G2. FKH-TF activity in G2 controls mammalian cell cycle termination, as interference with FKH transcriptional activation by disrupting PI(3)K/PKB downregulation, or by expressing a transcriptionally inactive FKH mutant, induces cell accumulation in G2/M, defective cytokinesis, and delayed transition from M to G1 of the cell cycle. We demonstrate that FKH-TFs regulate expression of mitotic genes such as *cyclin B* and *polo-like kinase (Plk)*. Our results support the important role of forkhead in the control of mammalian cell cycle completion, and suggest that efficient execution of the mitotic programme depends on downregulation of PI(3)K/PKB and consequent induction of FKH transcriptional activity.

To study the role of PI(3)K signalling on cell cycle progression, we compared the phenotypes of NIH3T3 cells expressing different interfering mutants of p85 and p110 PI(3)K^{6,7}. Expression of p65^{PI(3)K}, a mutant that enhances PI(3)K activation transiently⁶, accelerated transition through the different phases of the cell cycle (Fig. 1a). In contrast, expression of a constitutively active mutant, p110CAAX⁸, accelerated cell cycle entry² but slowed the transition between M and G1, thereby inducing cell accumulation in G2/M (Fig. 1a, b). This suggests that downregulation of PI(3)K activity after S phase may be required for execution of late cell cycle steps. In agreement, expression of higher p85α levels, which decreases endogenous PI(3)K activity⁶, reduced cell cycle entry but enhanced transition between M and G1 (Fig. 1a, b). Similar observations were made in T cells (not shown) and in fibroblasts expressing inducible PI(3)K mutants (see Supplementary Information Fig. 1).

We examined whether constitutive activation of PI(3)K affected progression through mitosis. Cells were stained for simultaneous visualization of the mitotic spindle, the actomyosin contractile ring formed during cytokinesis, and chromatin⁸ (Supplementary Information Fig. 2). This revealed that p110CAAX cells accumulated in telophase (Fig. 1c), with an increased proportion of binucleated cells (not shown), indicating defective cytokinesis.

To define the signalling pathways through which constitutive PI(3)K activation affects cytokinesis, we considered the PI(3)K effector PKB⁹. PKB activity was examined by western blot at different times after G0 release using anti-phospho-S473-PKB antibody⁵. PKB modification increased transiently after addition of serum and showed a second, minor activation peak in late G1

(17–19 h; Fig. 1d), which is required for G1/S transition³. PKB activation decreased thereafter and was low in cells arrested in G2, M and G0 (Fig. 1d). In contrast, PKB was highly phosphorylated at all phases of the cell cycle in p110CAAX cells (Fig. 1d). Stable cell lines expressing the constitutively active mutant Gag-PKB⁹ accumulated in G2/M and in telophase (Fig. 1e), and showed decelerated transition from M to G1 (not shown), resembling the behaviour of p110CAAX cells. This suggests that sustained activation of PI(3)K delays cytokinesis via PKB.

PKB phosphorylates FKH-TFs, inducing translocation to the cytosol and impairing nuclear transcriptional activity^{4,10–12}. FKH-TFs control mitosis in yeast^{13,14}; we thus considered that FKH-TFs may also be essential for mammalian mitosis, and that sustained PKB activation in p110CAAX cells may impair nuclear localization of FKH-TFs. We examined whether phosphorylation of endogen-

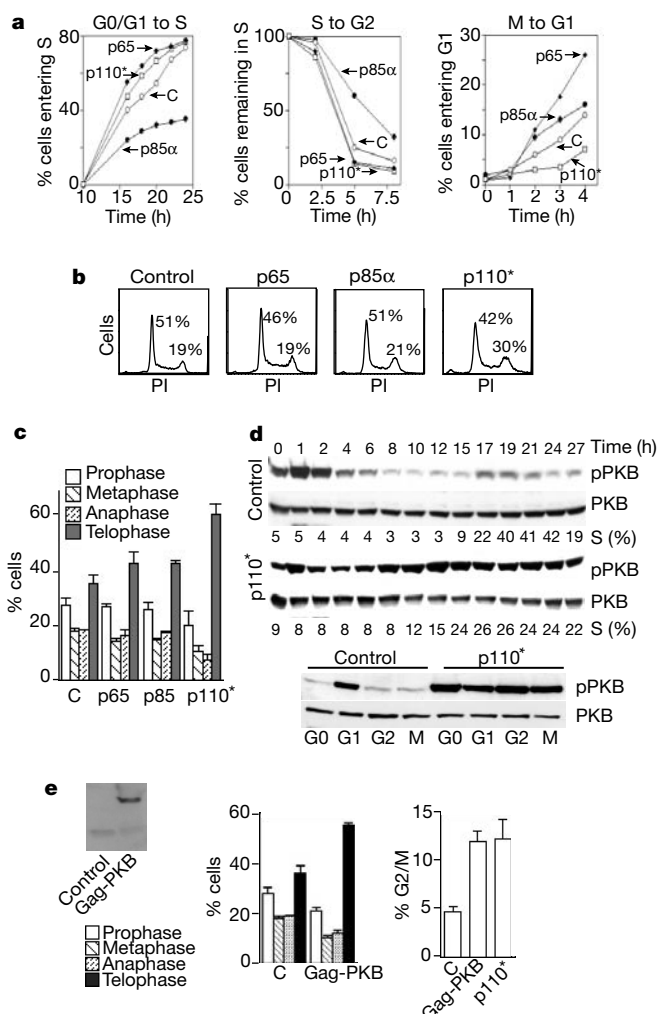


Figure 1 Sustained PI(3)K activation impairs cell cycle termination via Akt. **a**, Transition through different cell cycle phases of NIH3T3 cells expressing control vector (C), p65^{PI(3)K}, p85α and p110CAAX (p110*) (measured as detailed in Methods). **b**, Cell cycle profiles of the indicated cells in exponential growth. **c**, Percentage of G0/G1 and G2/M cells is shown. PI, propidium iodide. **d**, Percentage of cells at different stages of mitosis with respect to total number of cells in mitosis (100%). Mean ± s.d. of four experiments is shown. **e**, Western blot analysis of cells synchronously entering the cell cycle, using anti-PKB and anti-phospho-S473-PKB (pPKB) antibody (cells in S phase are indicated). Other samples were arrested in G0, G2, M, or arrested in G0 and incubated for 1 h with serum (G1). **e**, Western blot analysis of Gag-PKB expression in stable cell lines. Quantitative examination of cells at different mitosis stages as in **c** (left histogram) and percentage of cells in G2/M after 16 h incubation in medium with 0.1% serum (right histogram). Mean ± s.d.

ous FKHRL1, a member of the FKH-TF family¹¹, was regulated during the cell cycle. When control cells exited G0, both the amount and phosphorylation levels of FKHRL1 increased sharply, then decreased slowly as cells proceeded through G1, with a minor second phosphorylation peak approximately 19 h after addition of serum (Fig. 2a). FKHRL1 phosphorylation decreased thereafter and was low in cells that were arrested in G2, M and G0 (Fig. 2a). Although p110CAAX cells showed an increase in FKHRL1 protein levels in G1, these cells had a high phospho-FKHRL1 content at all times after addition of serum, which diminished moderately only when FKHRL1 protein levels decreased (Fig. 2a). Accordingly, endogenous FKHRL1 was localized in the nucleus in control cells

arrested in G0; it translocated to the cytosol in G1 and partially relocated to the nucleus in G2 (Fig. 2b). In contrast, FKHRL1 was always cytosolic in p110CAAX cells (Fig. 2b). p27 expression in G0 is induced by FKH-TFs⁵. The defective upregulation of p27 in p110CAAX cells arrested in G0 confirmed that these cells lack nuclear FKH-TF activity (Fig. 2c).

To determine whether the defect in FKH-TF was responsible for the G2/M delay in p110CAAX cells, we tested whether restoration of FKH-TF activity by overexpression of wild-type FKHRL1 or a PI(3)K/PKB-insensitive FKHRL1 mutant (FKHRL1A3)¹¹ corrected

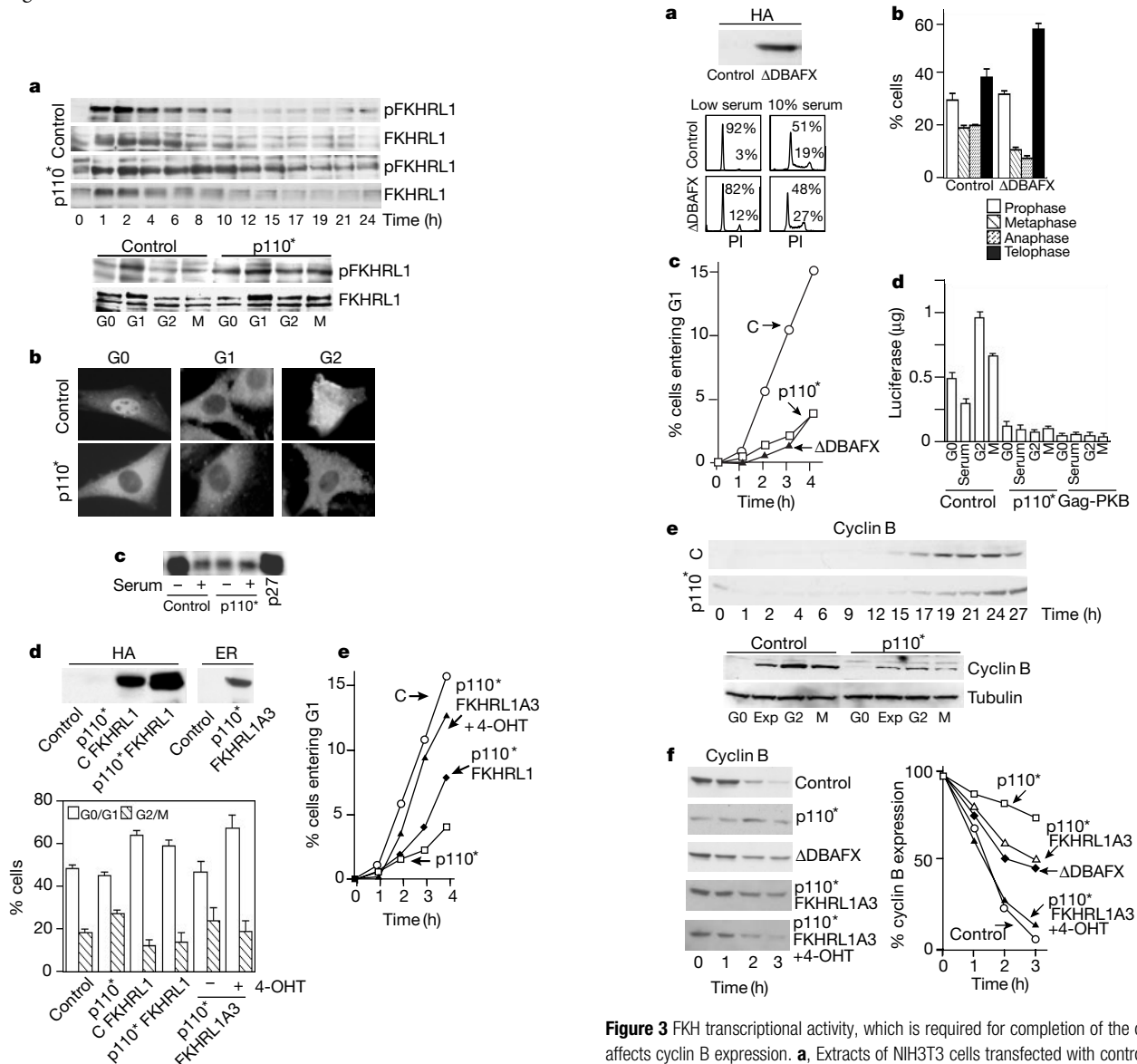


Figure 2 Constitutive activation of PI(3)K impairs endogenous FKHRL1 activity. **a**, Cells synchronized or arrested as in Fig. 1d were lysed; lysates were examined by western blot using anti-FKHRL1 and anti-phospho-T32-FKHRL1 antibody. **b**, Representative FKHRL1 immunofluorescence staining of control and p110CAAX cells arrested in G0 or G2, or arrested in G0 and treated with serum for 2 h (G1). **c**, p27 expression in G0-arrested or exponentially growing control and p110CAAX cells. Positive control of cells transfected with p27 is shown. **d**, Expression of stably transfected HA-FKHRL1 or ER-FKHRL1A3 in NIH3T3 or p110CAAX cells examined by western blot using anti-HA or anti-ER antibody, respectively. Percentage of cells in G0/G1 and G2/M is shown (bottom). For ER-FKHRL1A3 induction, cells were cultured for 16 h with 500 nM 4-OHT. **e**, Increase in the percentage of cells in G0/G1 after arrest at M phase with colcemid and release. ER-FKHRL1A3 was induced in G2 (see Methods).

Figure 3 FKH transcriptional activity, which is required for completion of the cell cycle, affects cyclin B expression. **a**, Extracts of NIH3T3 cells transfected with control or ΔDBAFX vectors were examined by western blot using anti-HA antibody. Cell cycle profiles of cells grown in medium with 10% or 0.1% serum are shown. Percentage of cells in G0/G1 and G2/M is also shown. PI, propidium iodide. **b**, Percentage of cells at different stages of mitosis (as in Fig. 1c). **c**, Increase in the percentage of cells in G0/G1 after arrest of M phase with colcemid and release. **d**, FKH reporter activity (arbitrary light units) in NIH3T3, p110CAAX and Gag-PKB cells arrested in G0, G2 or M, or arrested in G0 and incubated for 16 h in medium with 10% serum. Mean $\pm$ s.d. of four experiments is shown. **e**, Cyclin B expression by western blot in cells synchronously entering the cell cycle. Other samples were maintained in exponential growth (Exp) or arrested at different stages of the cell cycle. Tubulin expression is shown below. **f**, Cyclin B expression in the distinct cell lines arrested with colcemid and released for the indicated time periods. ER-FKHRL1A3 induction was as in Fig. 2e. Quantitative examination of cyclin B expression in the different cell lines compared to the amount of cyclin B in each line at time 0 (100%).

these cell cycle defects. We used an endoplasmic reticulum (ER)-inducible form for FKHL1A3 expression, as this mutant impairs cell survival¹¹. Stable expression of FKHL1 and induction of ER-FKHL1A3 increased the proportion of cells in G0/G1 as reported⁵ (Fig. 2d), and triggered activation of a FKHL1-reporter construct¹¹ (not shown). FKHL1 expression partially restored transition from M to G1 in p110CAAX cells (Fig. 2e). Moreover, ER-FKHL1A3 induction in G2 (Fig. 2e), but not in M phase (not shown), efficiently increased transition from M to G1. This suggests that

FKH transcriptional activation in G2 is required for completion of the cell cycle.

To further analyse whether FKH transcriptional activity is required for cytokinesis and exit of the M phase in mammalian cells, we disrupted endogenous FKH-TF action in NIH3T3 cells. Cells expressing a transcriptionally inactive mutant, Δ DBAFX⁵, accumulated in G2/M and in telophase (Fig. 3a, b), and exhibited defective transition from M to G1 (Fig. 3c). Interference with endogenous FKH activity by sustained PI(3)K/PKB activation or by expression of a transcription-defective FKH mutant delays cytokinesis and exit from the M phase. FKH-TF activity should then be induced normally in G2. Using an FKH-responsive *Fas ligand* promoter-reporter construct¹¹, we found that FKH-TF activity was low in cells that expressed p110CAAX or Gag-PKB (Fig. 3d), and decreased after addition of serum in control cells (Fig. 3d), as well as for cells treated with insulin or insulin-like growth factor I (refs 5, 11). Nonetheless, the highest FKH-TF activity levels were found in cells that were arrested at G2; remaining high in cells arrested at M phase (Fig. 3d). These results show that transcription mediated by FKH increases in G2 and is required for efficient cytokinesis and exit from M phase.

The phenotype of p110CAAX, Gag-PKB and Δ DBAFX cells, all with defects in FKH-TF activity, to some extent resembled that of yeast with FKH mutations, which exhibit defects in cell separation^{13,14}. In yeast, *FKH* genes control transcription of *CLB2* cluster genes, which regulate cyclin B expression as well as its eventual degradation^{13,14}. As cyclin B expression defects affect telophase and exit from the M phase^{15–17}, we compared cyclin B expression profiles in control and p110CAAX cells. Whereas cyclin B expression peaked 19–24 h after G0 release in control cells (maximum proportion of cells in G2/M), it accumulated slowly and was defectively degraded in p110CAAX cells (Fig. 3e). Sub-optimal cyclin B synthesis was also observed in p110CAAX cells arrested in G2 and M phase (Fig. 3e). Defective cyclin B degradation was confirmed by examining cyclin B levels in p110CAAX cells that were arrested in M phase and subsequently released (Fig. 3f). Suboptimal cyclin B degradation was also detected in Δ DBAFX cells (Fig. 3f). ER-FKHL1A3 induction in G2 restored normal cyclin B degradation kinetics in p110CAAX cells (Fig. 3f). FKH-TF thus regulates cyclin B synthesis and degradation.

We next examined whether mammalian FKH-TF, like yeast homologues^{13,14}, controls transcription of *cyclin B* and of genes that regulate cyclin B degradation. Among yeast FKH targets¹⁴, Plk modulates the ubiquitination of cyclin B, targeting it for degradation without affecting prior ubiquitination of anaphase inhibitors^{15,18,19}. We thus analysed whether FKH-TF regulates *cyclin B* and *Plk* expression in mammals. Plk protein levels were low in p110CAAX cells (Fig. 4a), and expression of either wild-type FKHL1 or FKHL1A3 (ref. 11) augmented cyclin B and Plk protein (Fig. 4b) as well as messenger RNA levels (not shown). In addition, the human *cyclin B* and *PLK* promoters^{20,21} contained sequence motifs similar to FKH-TF consensus binding sites^{11,14} (Fig. 4c), which were also found in the murine genes (not shown). Human *cyclin B* and *PLK* promoter fragments containing FKH-TF motifs were inserted upstream of a basal promoter that controlled luciferase expression. Co-transfection of these constructs with FKHL1 or FKHL1A3, but not with Δ DBAFX, markedly induced reporter expression in cells that were arrested in G2 (Fig. 4d). In contrast, FKH-TF did not activate the *MYC* promoter²² (Fig. 4d). Chromatin immunoprecipitation assays of human cells that were arrested in G2 showed that endogenous FKHL1 binds *in vivo* to the *cyclin B* and *PLK* promoter regions described above (Fig. 4e), but not to a control *cyclin D3* promoter region (Supplementary Information). FKH-TF thus regulates *cyclin B* and *Plk* expression by binding to their promoters. The defect in Plk expression may in turn affect cyclin B degradation^{15,18,19}. The combined action of defective Plk and cyclin B expression^{15–18,23}, and

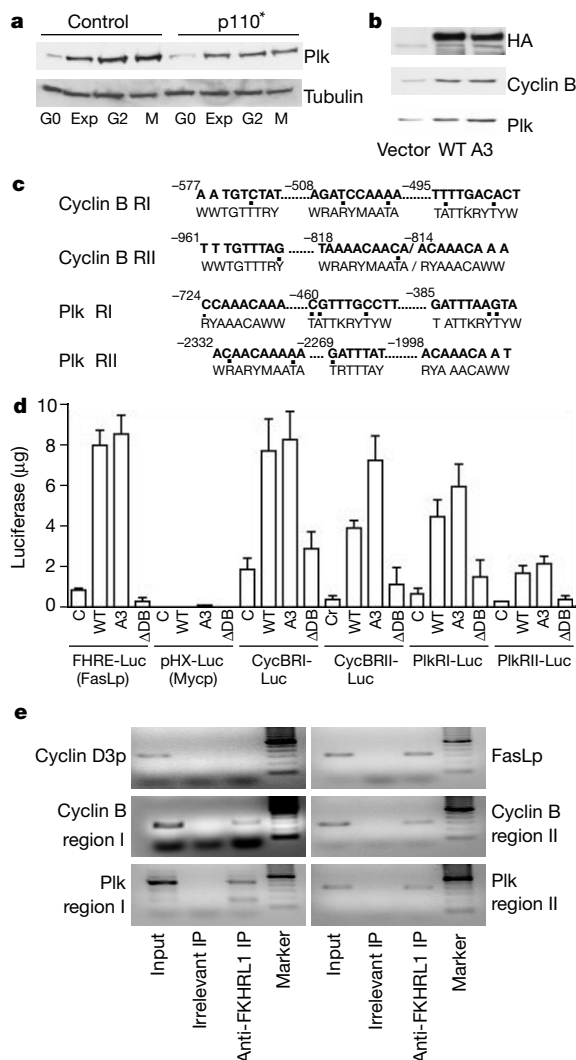


Figure 4 FKHL1 binds *cyclin B* and *Plk* promoters. **a**, Plk expression examined by western blot in NIH3T3 and p110CAAX cells arrested at different stages of the cell cycle as in Fig. 3e. **b**, NIH3T3 cells transiently transfected with control, HA-FKHL1 or HA-FKHL1A3 vectors were examined in western blot (indicated). **c**, Areas of the human *PLK* and *cyclin B* promoters containing motifs similar to FKH consensus: RYAAACAWW and WWTGTTTRY, sense and antisense FKH1 consensus; WRARYMAATA and TATTKRYTYW, sense and antisense *Xenopus* FKH consensus; TRTTTAY, antisense FKH human consensus (R = A/G, Y = T/C, W = A/T, M = A/C, K = T/G). Black boxes indicate mismatches. **d**, Luciferase reporter activity (light units) in NIH3T3 cells transfected with the indicated plasmids alone or in combination with HA-FKHL1 (WT), ER-FKHL1A3 (A3) or Δ DBAFX (Δ DB). Cells were arrested in G2 with etoposide before examination. **e**, FKHL1 was immunoprecipitated from chromatin extracts from HeLa cells arrested in G2, using normal rabbit serum or anti-FKHL1 antibody. Eluted DNA was subjected to PCR using oligonucleotides (Supplementary Information) flanking the *PLK* and *cyclin B* promoter regions I and II described in **c**. FasLp, *Fas ligand* promoter; Mycp, *MYC* promoter.

possibly of other mitotic genes, may contribute to the cytokinesis and M-phase exit defects in cells with impaired FKH-TF activity. Although cyclin B and Plk are required at mitosis stages before cytokinesis^{18,24}, both yeast^{13,14} and NIH3T3 cells with disruption of FKH-TF pathways seem to have sufficient residual cyclin B and Plk to transit through earlier mitotic stages. This supports the idea that other factors control *cyclin B* and *Plk* expression^{15,20}.

Completion of the mammalian cell cycle thus requires inactivation of the PI(3)K/PKB pathway and consequent FKH-TF activation. The mechanisms by which FKH genes regulate cytokinesis and transition from M to G1 include regulation of *cyclin B* and *Plk* transcription in G2. Future studies will aim to identify other mitotic genes regulated by FKH-TF in mammals, and to define whether, as in yeast, mammalian FKH-TF acts in concert with transcription factors containing the conserved eukaryotic MADS box motif^{13,14}. The current observations show a function for mammalian FKH-TF in the control of mitotic gene expression, which is conserved from yeast to mammals. □

Methods

Complementary DNAs, antibodies and transfections

pSG5-Gag-PKB, pMT2HA-ΔDBAFK and pcDNA3-ER-FKHRL1A3 were previously described^{5,9,25}. Wild-type pECE-HA-FKHRL1, pECE-HA-FKHRL1-A3 and the pGL3-based construct including the *Fas* ligand promoter region containing FKH consensus binding sites (FHRE-Luc reporter) were donated by M. E. Greenberg¹¹. The pGL3-based pHX-Luc vector, including the human *MYC* promoter, was a gift of K. Sugamura²². The human *cyclin B* and *PLK* promoter regions I and II were subcloned into pGL3 (Promega; Supplementary Information). Antibodies used were anti-haemagglutinin (HA) (Babco), anti-PKB, anti-RE (both Santa Cruz), anti-phospho-S473-PKB (Cell Signaling Tec), anti-FKHRL1, anti-phospho-FKHRL1 (T32) (both Upstate Biotechnology), anti-cyclin B1 (PharMingen), anti-Plk (Calbiochem) and anti-α-tubulin (Oncogene Research). p110CAAX, p85α and p65^{PI3K} NIH3T3 cells were previously described⁶⁷. Stable cell lines expressing Gag-PKB and the FKH-TF forms were obtained by transfection with the appropriate cDNA combined with pSVNeo or p-Pur (Clontech) using Tfx-50 (Promega). Transient transfections were performed using Lipofectamine Plus (Gibco-BRL).

Cell cycle analysis, arrest and transitions

We performed cell cycle analysis as described². Cells were synchronized in G0 by growth to confluence, maintained confluent for 48 h, then released by low-density replating (1:6) in 10% serum. For arrest at the G2 phase, cells were incubated for 20 h with 5 μM etoposide (Sigma), which yielded 40–50% of cells in G2. For M-phase arrest, cells were incubated for 10–16 h with 0.1 μg ml⁻¹ colcemid (Gibco), yielding approximately 70% of cells in M phase. To examine G1/S transition, cells were arrested in G0 and released by subconfluent plating in medium containing 10% serum and bromodeoxyuridine (BrdU; 10 mM) for different time periods, then stained with anti-BrdU-fluorescein isothiocyanate antibody and propidium iodide for cell cycle analysis, as described²⁶. For S-phase exit (S/G2 transition), cells growing exponentially were labelled with BrdU for 90 min. Under these conditions, most BrdU+ cells were in S phase (approximately 90%); the exit rate was monitored after BrdU deprivation by examining the percentage of BrdU+ cells remaining in S phase. To examine exit from the M phase, cell cultures that were synchronized at G0 were incubated for 14 h with serum and BrdU; under these conditions, most cells labelled with BrdU are in S phase. Cells were then deprived of BrdU and arrested with colcemid for 10 h, yielding 90–95% BrdU-labelled cells in G2/M phase. Exit from the cell cycle was monitored by tracking the appearance of BrdU+ cells in G0/G1 after colcemid deprivation (Fig. 1). Transition from M to G1 was also examined by omitting BrdU labelling, which yielded comparable results (Figs 3 and 4): cells were arrested in G0, incubated for 14 h in medium with 10% serum, then treated with colcemid for 10 h, yielding 90–95% cells in G2/M phase. Exit from the M phase was examined after the appearance of cells in G0/G1.

ER-FKHRL1A3 induction

For ER-FKHRL1A3 induction, cells were incubated with 500 nM 4-hydroxytamoxifen (4-OHT; Sigma). ER-FKHRL1A3 was induced as cells progressed through G2 by culturing cells during exponential growth without 4-OHT, followed by 6 h incubation in medium with 10% serum and colcemid (0.1 μg ml⁻¹). During exponential growth (in the presence of colcemid), 6 h is sufficient for most of the cells to finish S phase (not shown), and most cells are in G2 or entering G2. Cells were incubated for an additional 10 h in medium with 10% serum and colcemid (0.1 μg ml⁻¹) in 500 nM 4-OHT. Cells were washed and released in medium containing 4-OHT without colcemid. 4-OHT induction in G2-arrested cells treated with etoposide was avoided, as this arrest impairs progression of the cell cycle after etoposide deprivation. For ER-FKHRL1A3 induction in mitosis, we added 4-OHT after 16 h colcemid treatment.

Biochemistry, immunofluorescence, reporter assays

We performed western blot and luciferase assays as described^{6,11}. Light units were normalized using protein content and transfection efficiency, followed by co-transfecting

pEF-Bos-rCD2 (ref. 27). FKH immunofluorescence studies¹², cyclin B degradation analysis²⁸ and chromatin immunoprecipitation assays²⁹ were performed as described. After immunoprecipitation with anti-FKHRL1 antibody or normal rabbit serum, the DNA recovered was analysed by PCR using appropriate primers (Supplementary Information). We verified PCR products by sequencing.

Received 2 August; accepted 6 September 2001.

- Pardee, A. B. G1 events and regulation of cell proliferation. *Science* **246**, 603–608 (1989).
- Klippel, A. *et al.* Activation of phosphatidylinositol 3-kinase is sufficient for cell cycle entry and promotes cellular changes characteristic of oncogenic transformation. *Mol. Cell. Biol.* **18**, 5699–5711 (1998).
- Jones, S. M., Klinghoffer, R., Prestwich, G. D., Toker, A. & Kazanietz, A. PDGF induces an early and a late wave of PI 3-kinase activity, and only the late wave is required for progression through G1. *Curr. Biol.* **9**, 512–521 (1999).
- Kops, G. J. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
- Medema, R. H., Kops, G. J., Bos, J. L. & Burgering, B. M. AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27^{Kip1}. *Nature* **404**, 782–787 (2000).
- Jiménez, C. *et al.* Identification and characterization of a new oncogene derived from the regulatory subunit of phosphoinositide 3-kinase. *EMBO J.* **17**, 743–753 (1998).
- Jiménez, C. *et al.* Role of the PI3K regulatory subunit in the control of actin organization and cell migration. *J. Cell. Biol.* **151**, 249–262 (2000).
- Tatsumoto, T., Xie, X., Blumenthal, R., Okamoto, I. & Miki, T. Human ECT2 is an exchange factor for Rho GTPases, phosphorylated in G2/M phases, and involved in cytokinesis. *J. Cell. Biol.* **147**, 921–928 (1999).
- Burgering, B. M. & Coffey, P. J. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **376**, 599–602 (1995).
- del Peso, L., González, V. M., Hernández, R., Barr, F. G. & Núñez, G. Regulation of the forkhead transcription factor FKHR, but not the PAX3-FKHR fusion protein, by the serine/threonine kinase Akt. *Oncogene* **18**, 7328–7333 (1999).
- Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
- Jackson, J. G., Kreisberg, J. I., Koterba, A. P., Yee, D. & Brattain, M. G. Phosphorylation and nuclear exclusion of the forkhead transcription factor FKHR after epidermal growth factor treatment in human breast cancer cells. *Oncogene* **19**, 4574–4581 (2000).
- Pic, A. *et al.* The forkhead protein Fkh2 is a component of the yeast cell cycle transcription factor SFF. *EMBO J.* **19**, 3750–3761 (2000).
- Zhu, G. *et al.* Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* **406**, 90–94 (2000).
- Morgan, D. O. Regulation of the APC and the exit from mitosis. *Nature Cell Biol.* **1**, E47–E53 (1999).
- Murray, A. W., Solomon, M. J. & Kirschner, M. W. The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. *Nature* **339**, 280–286 (1989).
- Surana, U. *et al.* Destruction of the CDC28/CLB mitotic kinase is not required for the metaphase to anaphase transition in budding yeast. *EMBO J.* **12**, 1969–1978 (1993).
- Glover, D. M., Hagan, I. M. & Tavares, A. A. M. Polo-like kinases: a team that plays throughout mitosis. *Genes Dev.* **12**, 3777–3787 (1998).
- Kotani, S., Tanaka, H., Yasuda, H. & Todokoro, K. Regulation of APC activity by phosphorylation and regulatory factors. *J. Cell. Biol.* **146**, 791–800 (1999).
- Cogswell, J. P., Godlevski, M. M., Bonham, M., Bisi, J. & Babiss, L. Upstream stimulatory factor regulates expression of the cell cycle-dependent cyclin B1 promoter. *Mol. Cell. Biol.* **15**, 2782–2790 (1995).
- Bräuninger, A., Strehhardt, K. & Rübsamen-Waigmann, H. Identification and functional characterization of the human and murine polo-like kinase (Plk) promoter. *Oncogene* **11**, 1793–1800 (1995).
- Takeshita, T. *et al.* STAM, signal transducing adaptor molecule, is associated with Janus kinases and involved in signaling for cell growth and c-myc induction. *Immunity* **6**, 449–457 (1997).
- Carmena, M. *et al.* *Drosophila* polo kinase is required for cytokinesis. *J. Cell. Biol.* **143**, 659–671 (1998).
- Zachariae, W. Progression into and out of mitosis. *Curr. Opin. Cell Biol.* **11**, 708–716 (1999).
- Dijkers, P. F., Medema, R. H., Lammers, J. W., Koenderman, L. & Coffey, P. J. Expression of the pro-apoptotic Bcl-2 family member Bim is regulated by the forkhead transcription factor FKHR-L1. *Curr. Biol.* **10**, 1201–1204 (2000).
- Borlado, L. R. *et al.* Increased phosphoinositide 3-kinase activity induces a lymphoproliferative disorder and contributes to tumor generation *in vivo*. *FASEB J.* **14**, 895–903 (2000).
- Reif, K., Nobes, C. D., Thomas, G., Hall, A. & Cantrell, D. A. Phosphatidylinositol 3-kinase signals activate a selective subset of Rac/Rho-dependent effector pathways. *Curr. Biol.* **6**, 1445–1455 (1996).
- Kotani, S. *et al.* PKA and MPF-activated polo-like kinase regulate anaphase-promoting complex activity and mitosis progression. *Mol. Cell* **1**, 371–380 (1998).
- Boyd, K. E., Wells, J., Gutman, J., Bartley, S. M. & Farnham, P. J. c-Myc target gene specificity is determined by a post-DNA-binding mechanism. *Proc. Natl Acad. Sci. USA* **95**, 13887–13892 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Gonzalo for advice in chromatin immunoprecipitation assays and M. Serrano, R. Medema, I. Mérida, M. Torres, D. Jones, V. Calvo and C. Mark for critical reading of the manuscript. This work was supported by grants from the Community of Madrid, DGCyT (Dirección General de Ciencia y Tecnología) and the Pharmacia Corporation. The Department of Immunology and Oncology was founded and is supported by the Spanish Council for Scientific Research (CSIC) and by the Pharmacia Corporation.

Correspondence and requests should be addressed to A.C.C. (e-mail: acarrera@cnb.uam.es).

The bacteriophage $\phi 29$ portal motor can package DNA against a large internal force

Douglas E. Smith^{†,‡}, Sander J. Tans^{†,‡}, Steven B. Smith[‡], Shelley Grimes[§], Dwight L. Anderson[§] & Carlos Bustamante^{†,‡,¶}

[†]Department of Physics, [‡]Department of Molecular and Cell Biology, [§]Howard Hughes Medical Institute, [¶]Physical Biosciences Division of Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA
[§]Departments of Microbiology and Oral Science, University of Minnesota, Minneapolis, Minnesota 55455, USA

* These authors contributed equally to this work

As part of the viral infection cycle, viruses must package their newly replicated genomes for delivery to other host cells. Bacteriophage $\phi 29$ packages its 6.6- μm long, double-stranded DNA into a $42 \times 54 \text{ nm}$ capsid¹ by means of a portal complex that hydrolyses ATP². This process is remarkable because entropic, electrostatic and bending energies of the DNA must be overcome to package the DNA to near-crystalline density. Here we use optical tweezers to pull on single DNA molecules as they are packaged, thus demonstrating that the portal complex is a force-generating motor. This motor can work against loads of up to 57 pN on average, making it one of the strongest molecular motors reported to date. Movements of over 5 μm are observed, indicating high processivity. Pauses and slips also occur, particularly at higher forces. We establish the force-velocity relationship of the motor and find that the rate-limiting step of the motor's cycle is force dependent even at low loads. Notably, the packaging rate decreases as the prohead is filled, indicating that an internal force builds up to $\sim 50 \text{ pN}$ owing to DNA confinement. Our data suggest that this force may be available for initiating the ejection of the DNA from the capsid during infection.

The *Bacillus subtilis* phage $\phi 29$ is an excellent model system for studying viral assembly³. The 19.3-kilobase (kb) genome with its terminal proteins, gp3, is packaged into a preformed capsid (prohead) in an efficient *in vitro* assay^{4,5}. The DNA packaging motor lies at a unique portal vertex of the prohead^{1,6,7} and contains: (1) the head-tail connector (a dodecamer of gp10), the structure of which has been solved recently by X-ray crystallography⁶; (2) the portion of the prohead shell that surrounds the connector; (3) a ring of 174-base prohead RNAs (pRNA), which surrounds the protruding narrow end of the connector; and (4) a multimer of gp16, an ATPase that first binds DNA-gp3 and then assembles onto the connector/pRNA complex prior to packaging. It has been proposed that this portal complex represents a new class of rotary molecular motors, which couple rotation to DNA translocation⁶.

Previously, bacteriophage DNA packaging has been studied in bulk reactions using DNase protection assays^{4,5}. Here we present an assay in which force-measuring laser tweezers are used to follow packaging activity of a single complex in real time. Stalled, partly prepackaged complexes are attached to a polystyrene microsphere by means of the unpackaged end of the DNA (see Methods). This microsphere is captured in the optical trap and brought into contact with a second bead that is held by a pipette. This bead is coated with antibodies against the phage, so a stable tether is formed between the two beads (Fig. 1a). In the absence of ATP, the tether displays the elasticity expected for a single DNA molecule⁸. Shortly after addition of ATP, the two microspheres move closer together, indicating packaging activity. Control experiments confirm this conclusion: proheads bound to antibody-coated microspheres efficiently package biotinylated $\phi 29$ DNA in a bulk assay; no movement is observed in the measurements using optical tweezers when the proheads are

omitted and the DNA is directly attached to the beads; and packaging could be reversibly stalled by a non-hydrolysable ATP analogue ($\gamma\text{S-ATP}$).

The first set of measurements was carried out in 'constant force feedback' mode, in which the microsphere distance is adjusted by feedback to maintain DNA tension at a preset value of 5 pN (Fig. 1a). Figure 1b shows plots of the tether length versus time for a $\phi 29$ - λ DNA construct with a length of 1.8 times the $\phi 29$ genome (see Methods). In all measurements a saturating concentration of ATP was used (0.5 mM). Packaging is highly efficient: in over 95% of the measurements, movements of several micrometres could be followed. Packaging dynamics are recorded with more detail, accuracy and time resolution than has previously been possible in bulk studies. Accounting for the prepackaging, we find that it takes 5.5 min on average to package a length of DNA equal to the $\phi 29$ genome, with a standard deviation from complex to complex of 0.8 min. A single complex also shows packaging rate fluctuations of $\sim 20 \text{ bp s}^{-1}$ r.m.s. in a 1 Hz bandwidth (Fig. 1c), which are approximately five times larger than the measurement noise of $\sim 4 \text{ bp s}^{-1}$ r.m.s., determined when the DNA alone is directly attached between two beads (see Supplementary Information Fig. I). Pauses in movement of variable duration (see Fig. 1b, inset) and slips (see below) are also clearly evident.

On average there are 3.1 pauses per micrometre of DNA packaged ($\sigma = 3.5$, 43 complexes) and these pauses have a mean duration of 4.0 s ($\sigma = 5.1 \text{ s}$, $n = 425$). At higher capsid fillings the pausing frequency increases (correlation coefficient⁹ of plots of pausing frequency versus capsid filling ~ 0.8), although the pause duration is not strongly correlated with filling (correlation coefficient of plots of pause duration versus capsid filling ~ 0.01). Details of the pause analysis are presented in Supplementary Information Figs II and III. During a pause, the DNA tether length remains constant despite the applied force, indicating that the motor stays engaged with the DNA. During packaging, however, some complexes occasionally exhibit abrupt increases in DNA tether length (see Fig. 2, inset). Apparently the motor loses grip on the DNA but can grasp the molecule again and resume packaging immediately. The mean length of a slip is 44 bp, with a large standard deviation of $\sigma = 103 \text{ bp}$ ($n = 99$). We have not found any evidence that slips occur at specific locations along the DNA, but they do occur more often at higher applied forces (see Supplementary Information Fig. IV). The motor is highly processive, in the sense that for the ensemble of records the length of DNA packaged divided by the number of slips is large ($\sim 7 \mu\text{m}$, under a load of 5 pN).

In Fig. 1c the packaging rate is plotted against the percentage of the genome that has been packaged. These measurements show a marked reduction in packaging rate beginning when $\sim 50\%$ of the genome is packaged. Initially the rate is $\sim 100 \text{ bp s}^{-1}$ and it gradually drops to zero as the capsid fills up and the motor stalls. Note that the proheads are able to package up to $\sim 5\%$ more DNA than the normal $\phi 29$ genome length. These observations suggest that the rate decrease results from the build up of an internal pressure due to DNA confinement, which exerts an opposing force and slows the motor.

Within such a model, this internal force generated inside the capsid could be quantified if the effect of force on the velocity of the motor were known. The force dependence of the motor was investigated using a 'no feedback' mode where the positions of the trap and pipette are fixed (see Fig. 1a). These measurements were done using $\phi 29$ -NcoI-A DNA ($0.8\times$ the $\phi 29$ genome; see Methods). In this mode the tension in the molecule increases as the motor reels in the DNA (Fig. 2a), and the bead is displaced from the centre of the trap. This force acts directly on the motor and at a certain point causes it to stall. These data allow us to determine how the packaging rate depends on applied force. Three such force-velocity ($F-v$) traces for different complexes are shown in Fig. 3a. The curve

shapes are similar, although the initial packaging rate and the stall force vary.

The mean F - v behaviour is obtained by first normalizing a number of individual curves to the mean stall force and mean initial velocity, and then averaging them together. Mean F - v curves were obtained at two points: when one-third and two-thirds of the genome is packaged (Fig. 3b, solid red and blue lines, respectively). Figure 1c suggests that in the case of one-third of the genome being packaged, no significant internal force opposes the motor because the velocity has not yet been reduced by DNA filling, although the motor does exhibit sensitivity to external force even at these low fillings (Fig. 3b). F - v data in the two-thirds case, however, suggest that an internal force of ~ 14 pN is acting on the motor, as it takes ~ 14 pN less external force to stall the motor in this case (Fig. 3b). Here we rely on the supposition that the internal and external forces

add together to act on the motor. This supposition is supported by the good overlap of the F - v relationships over their entire range when the data for two-thirds genome packaging is shifted by +14 pN along the force axis. In this way, the inherent F - v relationship of the motor in the absence of any internal forces can be obtained.

The F - v curve in Fig. 3b shows that the packaging rate decreases even for small forces, indicating that the rate-limiting step in the motor's mechano-chemical cycle is force-dependent and therefore involves a mechanical displacement. Such behaviour is different from that reported for RNA polymerase¹⁰, but is similar to that found in the bacterial flagellar motor¹¹. Within a Kramer's model¹² of thermal activation over a single reaction barrier under the influence of force, our data imply that the rate-limiting step produces a conformational change of only ~ 0.1 nm (see Supple-

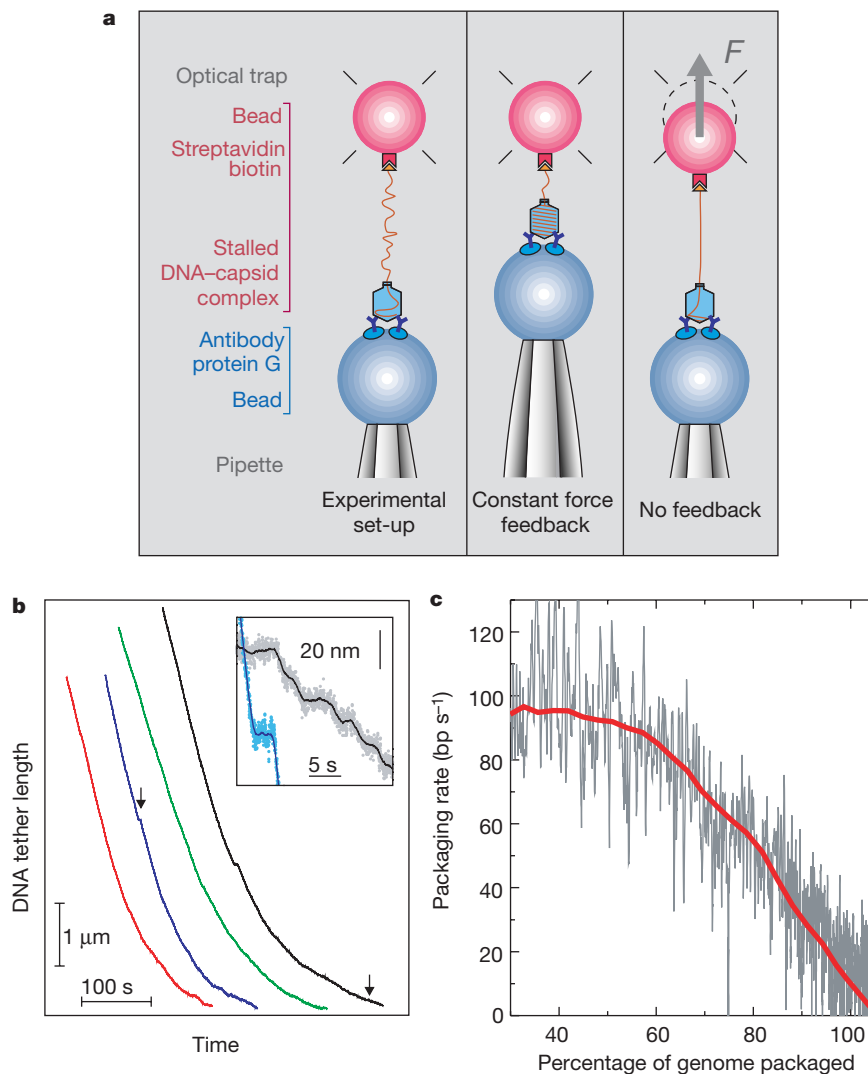


Figure 1 Set-up and initial results. **a**, Diagrams showing the experimental set-up at the start of a measurement (left), constant force feedback mode (middle) and no feedback mode (right) measurements. A single $\phi 29$ packaging complex is tethered between two microspheres. Optical tweezers are used to trap one microsphere and measure the forces acting on it, while the other microsphere is held by a micropipette. To insure measurement on only one complex, the density of complexes on the microsphere is adjusted so that only about one out of five–ten microspheres yielded hook-ups. Such attachments break in one discrete step as the force is increased, indicating that only one DNA molecule carries the load. **b**, DNA tether length against time for four different complexes with a constant force of ~ 5 pN using a 34.4-kb $\phi 29$ - λ DNA construct. Inset, increased detail of regions,

indicated by arrows, showing pauses (curves have been shifted). The solid lines are a 100-point average of the raw data. **c**, Packaging rate against the amount of DNA packaged, relative to the original 19.3-kb $\phi 29$ genome. Grey line, trace for a single complex (derived from black trace in **b**). Rates were obtained by linear fitting in a 1.5-s sliding window. The red line is an average of eight such measurements. Large pauses (velocity drops >30 bp s^{-1} below local average) were removed, and the curves were horizontally shifted to account for differences in microsphere attachment points. The red line was smoothed using a 200-nm sliding window. The standard deviation for the ensemble of measurements varies from ~ 20 bp s^{-1} at the beginning, down to ~ 10 bp s^{-1} at the end.

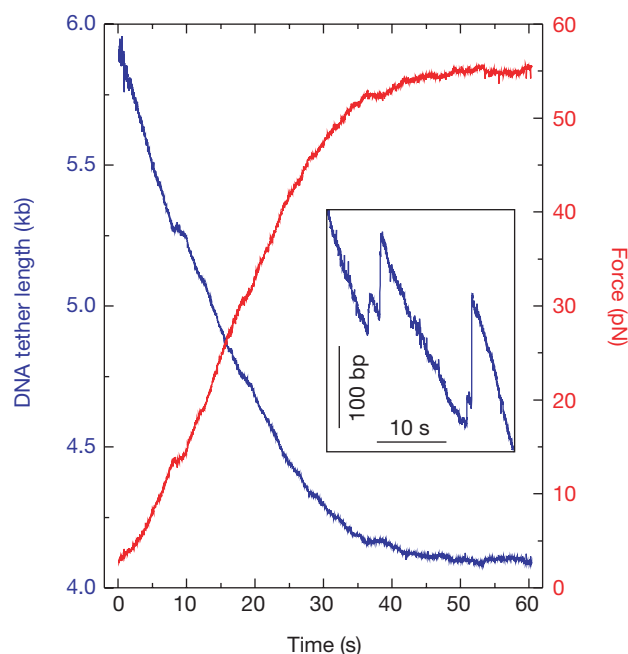


Figure 2 Measurements of packaging in no feedback mode. The tether shortens (blue line) and the force increases (red line) as the portal motor pulls in the DNA. In this example, the motor stalled at a force of ~ 55 pN. Although in many measurements the linkage broke before stalling was observed, similar maximum forces could be projected from

these curves. We suspect that such breakage events occur at the antibody–prohead connection. Inset, spontaneous slipping events, where the DNA comes out of the capsid, were accompanied with abrupt decreases in force.

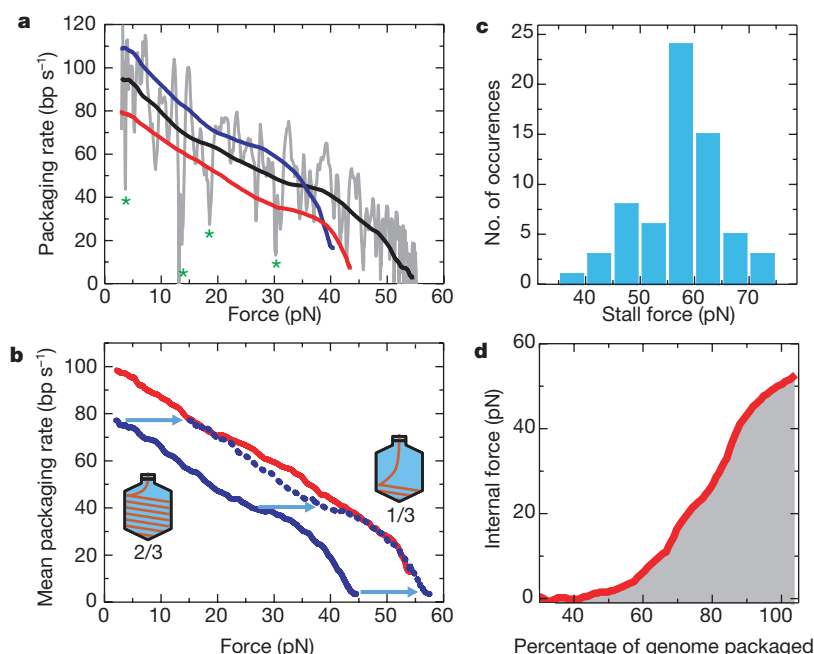


Figure 3 Force–velocity (F – v) analysis. **a**, The packaging rate for a single complex (grey line) was determined by linear fitting of the data in Fig. 2 in a 0.7-s sliding window. The black line is obtained by editing out large pauses (asterisks indicate where velocity drops >30 bp s^{-1} below local average) and smoothing (50-point sliding window). These long pauses were removed so as not to perturb the general trend of the F – v behaviour. The red and blue lines are data from two other complexes. **b**, External force against velocity curves when about one-third (red line) and about two-thirds (blue line) of the genome is packaged. Curves were obtained from averaging 14 and 8 individual traces, respectively. If, in the case of two-thirds of the genome being packaged, an additional 14 pN of internal force were acting on the motor (see text), the dashed blue line would show the expected behaviour. The red line and the dashed blue lines would then represent the inherent (total) F – v curve for the motor. **c**, Stall force measured for 65 individual complexes indicates an average stall force of ~ 57 pN. Stall force refers to the total force (external force plus, in the case of two-thirds of the genome being packaged, the inferred

internal force of 14 pN), which is needed to stop further packaging. Clear plateaux (slope <5 bp s^{-1} for >5 s) in the force against time plots were sometimes observed (for example, in Fig. 2). The average force over this plateau is taken as the stall force. In some cases the packaging rate dropped by more than 80% (to below <20 bp s^{-1}), but the tether broke before a full plateau was reached. In these cases we extrapolated to obtain an estimate of the stall force, as the shapes of truncated curves were similar to those that fully stalled. DNA over-stretching was not observed because linkages always broke at external forces lower than 65 pN. However, stall forces above 65 pN could be determined in cases where internal force added to the total force. **d**, Internal force against percentage genome packaged. This plot is obtained by relating the packaging rate, as obtained from the rate against percentage of genome packaged curve (Fig. 1c), to the total force required to produce the same packaging rate, as given by the rate against force curve (**b**, average of red and dashed blue lines). To obtain the internal force we subtracted the 5 pN of external force that is present in the experiment of Fig. 1c.

mentary Information Fig. V). This value is much smaller than the net movement of ~ 0.68 nm (2 bp) per ATP inferred from bulk studies². Therefore, to the extent that this simple model applies, the rate-limiting step is sensitive to force, but is not the main translocation step. Note that the velocity begins to decrease more sharply with force at ~ 45 pN, suggesting that a second step in the mechanochemical cycle of the motor, associated with a larger movement, becomes rate-limiting at these higher forces.

Figure 3c shows a histogram of the total force needed to stall the motor. The stall force ranges between 40 and 70 pN and has an average force of ~ 57 pN, making $\phi 29$ one of the strongest molecular motors reported to date. This stall force is approximately eight times higher than that reported for conventional kinesin^{13,14} or skeletal muscle myosin II motors¹⁵, and approximately two times higher than that for RNA polymerase¹⁰. By multiplying the average stall force (57 pN) with the distance moved per ATP² (~ 0.68 nm), we obtain a work done per ATP of ~ 39 pN nm. Assuming the free energy of ATP hydrolysis in our buffer¹⁶ is ~ 120 pN nm, then the energy conversion efficiency is $\sim 30\%$. This value falls within the range of efficiencies reported for myosin¹⁵, kinesin^{13,14} and RNA polymerase motors¹⁰. Additionally, a maximum force of 70 pN implies that the motor's step size must be less than 5 bp (120 pN nm/ 70 pN ≈ 1.7 nm).

Combining the rate dependence on external force (Fig. 3b) and on the fraction of DNA packaged (Fig. 1c), we can make a quantitative estimate of the build-up of internal force as the prohead is filled. Figure 3d shows that when the full genome has been packaged, this internal force reaches a large value of about 50 pN. In many phages it is believed that the DNA is ordered as a spool in which the strands are packaged in a hexagonal lattice. By dividing the measured force by the hexagonal cell surface area, as determined by X-ray diffraction studies¹⁷, one may obtain a rough estimate of a pressure of 6 MPa inside the phage capsid, assuming that there is no significant energy dissipation in packaging. An osmotic pressure of a similar order (~ 1 MPa) has been reported to be necessary to condense DNA in solution to a density similar to that in the phage capsid¹⁸. If 6 MPa of pressure is transmitted to the capsid shell, the shell (thickness ~ 1.6 nm¹) must have a tensile strength of at least 100 MPa, a value similar to the bulk tensile strength for a typical aluminum alloy.

As seen in Fig. 3d, the force starts building only after about half of the $\phi 29$ genome is already packaged, a trend that has recently been seen also in simulations of DNA packaging into phage capsids (J. Kindt, S. Tzllil, A. Ben-Shaul and W. M. Gelbart, personal communication). Our finding suggests that the DNA does not immediately adopt its final condensed state inside the capsid, but instead is progressively compressed. Furthermore, both the pause frequency and the internal pressure increase with DNA filling, indicating that internal pressure may have a function in inducing pauses.

The compression of DNA in phage capsids has long been considered an intriguing phenomenon because it is thought to involve large energetic penalties^{19,20}. An estimate of the total work done in packaging the $\phi 29$ genome is obtained by integrating the force curve in Fig. 3d, which yields 7.5×10^{-17} J (or 2×10^4 kT). This value can be compared with theoretical predictions for the equilibrium free energy change of the DNA. One model¹⁹ gives 4×10^4 kT, whereas a preliminary estimate (T. Odijk, personal communication) based on another model²⁰ yields $\sim 5.6 \times 10^3$ kT. These predictions, which include electrostatic, bending and entropic penalties, but no dissipative effects, are of the same order of magnitude as our estimate of the total work done. Thus it is possible that a large fraction of the work done by the motor is reversible work, and that dissipation is not dominant. Moreover, the observations of rapid slipping of the DNA out of the capsid on a timescale below 0.014 s with no evidence of slower relaxation also suggest that there is little viscous resistance to DNA movement within the capsid.

Whereas it has been argued that T7 phage relies on a motor for DNA injection into the host²¹, other phages are known to passively eject their genome when exposed to the bacterial membrane receptor^{17,22}. Our results lend support to a model for $\phi 29$ in which internal pressure provides the driving force for DNA injection into the host cell for the first half of the injection process. □

Methods

Stalled packaging complexes

Components of the $\phi 29$ *in vitro* packaging assay, including proheads, DNA-gp3 and gp16, were purified as described previously²³. The 19.3-kb $\phi 29$ DNA-gp3 (6.6 μ m in length) was cut with *Nco*I to yield a 15-kb left end ($\phi 29$ -*Nco*I-A DNA) and a 4.3-kb right end. The left end is preferentially packaged²⁴. These fragments were biotinylated using the Klenow fragment of DNA polymerase I (exo⁻ mutant; New England Biolabs) to incorporate biotin-14-dATP and biotin-14-dCTP (Gibco). We constructed the longer 34.4-kb $\phi 29$ - λ DNA by ligating together the $\phi 29$ and λ *Nco*I-A fragments. Before cutting, the cos sites (cohesive ends) of λ -DNA were biotinylated as described above. To initiate a packaging reaction we added 4 μ l of 0.25 mM ATP (Roche) to a 13.5 μ l solution of buffer A (50 mM Tris-HCl buffer (pH 7.8), 50 mM NaCl, 5 mM MgCl₂), 0.1 μ g of biotinylated DNA-gp3, 0.5 μ g gp16 and 10 μ g $\phi 29$ proheads. This mixture is incubated for approximately 30 s, and then the packaging is stalled by adding 4 μ l of 2.5 mM γ S-ATP (Roche), such that the genome is partly prepackaged. These complexes are stable for >10 h and are reactivated during the measurement by exposing them to buffer A plus 0.5 mM ATP, 5 μ M ADP, 5 μ M NaH₂PO₄ (source of phosphate ions) and 50 μ g ml⁻¹ BSA. This ATP concentration is saturating (increasing it by a factor of ten does not significantly increase the packaging rate, that is, the reaction is not limited by diffusion).

Microsphere preparation

Polystyrene microspheres that were coated with streptavidin (2.2 μ m diameter, 0.5% w/v; SpheroTech) were washed twice and pre-blocked for 5 min with 5 mg ml⁻¹ BSA in buffer A. Stalled complexes were attached to the microspheres (by means of the biotinylated DNA) by adding approximately 0.3 μ l of the stalled reaction mixture and 1 μ l RNase inhibitor (SUPERase-IN, Ambion) to 20 μ l of microspheres, and incubating them for 15–20 min. We then diluted the microspheres in 1 ml buffer A. Polystyrene microspheres coated with protein G (2.8 μ m diameter, 5% w/v; SpheroTech) were washed twice in phosphate-buffered saline and incubated for >20 min with a $\sim 1/10$ dilution of rabbit antisera prepared against $\phi 29$. We added 5 μ l of microspheres (washed five times) and 1 μ l of RNase inhibitor to 1 ml of buffer A.

Optical tweezers

We used a dual-beam optical tweezers apparatus similar to that described previously⁸. The measured trap stiffness is ~ 0.12 pN nm⁻¹ and the force is recorded at a rate of ~ 70 Hz. The displacements of the trapped microsphere were derived from the measured forces. The pulling angle could be varied over a 120° range but did not seem to affect the packaging rate. The pipette position is recorded using an optical lever having a resolution of ~ 1 nm. The contour length of the DNA (in bp) was determined from the measured force and end-to-end distance, and through using the worm-like chain model assuming a persistence length of 53 nm, a stretch modulus of 1,200 pN and a distance per bp of 0.34 nm²⁵.

Received 20 March; accepted 24 August 2001.

1. Tao, Y. *et al.* Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95**, 431–437 (1998).
2. Guo, P., Peterson, C. & Anderson, D. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage $\phi 29$. *J. Mol. Biol.* **197**, 229–236 (1987).
3. Anderson, D. & Reilly, B. in *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics* (eds Sonenshein, A., Hoch, J. A. & Losick, R.) 859–867 (American Society for Microbiology, Washington DC, 1993).
4. Guo, P., Grimes, S. & Anderson, D. A defined system for *in vitro* packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage $\phi 29$. *Proc. Natl Acad. Sci. USA* **83**, 3505–3509 (1986).
5. Grimes, S. & Anderson, D. *In vitro* packaging of bacteriophage $\phi 29$ DNA restriction fragments and the role of the terminal protein gp3. *J. Mol. Biol.* **209**, 91–100 (1989).
6. Simpson, A. A. *et al.* Structure of the bacteriophage $\phi 29$ DNA packaging motor. *Nature* **408**, 745–750 (2000).
7. Ibarra, B. *et al.* Topology of the components of the DNA packaging machinery in the phage $\phi 29$ prohead. *J. Mol. Biol.* **298**, 807–815 (2000).
8. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
9. Barlow, R. *Statistics: a Guide to the Use of Statistical Methods in the Physical Sciences* 15–16 (Wiley, Chichester, 1989).
10. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
11. Berg, H. C. & Turner, L. Torque generated by the flagellar motor of *Escherichia coli*. *Biophys. J.* **65**, 2201–2216 (1993).
12. Kramers, H. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
13. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77**, 773–784 (1994).

14. Coppin, C. M., Pierce, D. W., Hsu, L. & Vale, R. D. The load dependence of kinesin's mechanical cycle. *Proc. Natl Acad. Sci. USA* **94**, 8539–8544 (1997).
15. Finer, J. T., Simmons, R. M. & Spudis, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
16. Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Principles of Biochemistry* 375 (Worth, New York, 1993).
17. Earnshaw, W. C. & Casjens, S. R. DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**, 319–331 (1980).
18. Rau, D. C., Lee, B. & Parsegian, V. A. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. *Proc. Natl Acad. Sci. USA* **81**, 2621–2625 (1984).
19. Riemer, S. C. & Bloomfield, V. A. Packaging of DNA in bacteriophage heads: some considerations on energetics. *Biopolymers* **17**, 785–794 (1978).
20. Odijk, T. Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. *Biophys. J.* **75**, 1223–1227 (1998).
21. Garcia, L. R. & Molineux, I. J. Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli*. *J. Bacteriol.* **178**, 6921–6929 (1996).
22. Novick, S. L. & Baladeschwieler, J. D. Fluorescence measurement of the kinetics of DNA injection by bacteriophage lambda into liposomes. *Biochemistry* **27**, 7919–7924 (1988).
23. Grimes, S. & Anderson, D. The bacteriophage ϕ 29 packaging proteins supercoil the DNA ends. *J. Mol. Biol.* **266**, 901–914 (1997).
24. Bjornsti, M. A., Reilly, B. E. & Anderson, D. L. Morphogenesis of bacteriophage ϕ 29 of *Bacillus subtilis*: oriented and quantized *in vitro* packaging of DNA-gp3. *J. Virol.* **45**, 383–396 (1983).
25. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl Acad. Sci. USA* **94**, 6185–6190 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank W. M. Gelbart, P. Jardine, T. Odijk, V. Bloomfield, D. Frenkel, C. Varga, A. Mehta and M. Young for comments. This research was supported in part by grants from the NIH, DOE and NSF. D.E.S. and S.J.T. are supported by a grant from the Packard Foundation. S.J.T. is supported by the Netherlands Organization for Scientific Research (NWO).

Correspondence and requests for materials should be addressed to C.B. (e-mail: carlos@alice.berkeley.edu).

Uracil-DNA glycosylase acts by substrate autocatalysis

Aaron R. Dinner^{*†}, G. Michael Blackburn[‡] & Martin Karplus^{*§}

^{*}Central Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QH, UK

[‡]Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK

[§]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

^{||}Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 4 Rue Blaise Pascal, 67000 Strasbourg, France

In humans, uracil appears in DNA at the rate of several hundred bases per cell each day as a result of misincorporation of deoxyuridine (dU) or deamination of cytosine. Four enzymes that catalyse the hydrolysis of the glycosylic bond of dU in DNA to yield an apyrimidinic site as the first step in base excision repair have been identified in the human genome¹. The most efficient and well characterized of these uracil-DNA glycosylases is UDG (also known as UNG and present in almost all known organisms)², which excises U from single- or double-stranded DNA and is associated with DNA replication forks³. We used a hybrid quantum-mechanical/molecular-mechanical (QM/MM) approach⁴ to determine the mechanism of catalysis by UDG. In contrast to the concerted associative mechanism proposed initially^{5–10}, we show here that the reaction proceeds in a stepwise dissociative manner^{11,12}. Cleavage of the glycosylic bond yields an

intermediate comprising an oxocarbenium cation and a uracilate anion. Subsequent attack by a water molecule and transfer of a proton to D145 result in the products. Surprisingly, the primary contribution to lowering the activation energy comes from the substrate, rather than from the enzyme. This 'autocatalysis' derives from the burial and positioning of four phosphate groups that stabilize the rate-determining transition state. The importance of these phosphates explains the residual activity observed for mutants that lack key residues^{6–9}. A corresponding catalytic mechanism could apply to the DNA glycosylases TDG and SMUG1, which belong to the same structural superfamily as UDG^{13,14}.

In the QM/MM simulations, the reactive region was treated with the AM1 method¹⁵ (and higher-level techniques, as described in Methods), while the surrounding non-reactive region was modelled classically with the CHARMM force field^{4,16}. Because the enzyme–substrate complex (described below) has many charged groups, solvent shielding is essential, and we adapted a procedure developed for MM free energy simulations¹⁷ to QM/MM calculations (see Methods). The charges of exposed groups were scaled to avoid distortion of the structures during the simulations, and continuum electrostatics methods were applied subsequently to obtain energies that include the correct solvent-shielding effects.

The starting point for the simulations was the structure of an enzyme–inhibitor complex at 1.8 Å resolution¹¹. This complex contains the non-hydrolysable C-glycoside deoxypseudouridine (dΨU) in place of the dU substrate (see Supplementary Information Fig. 1). The calculations include the entire recombinant protein (223 residues, indexed according to their positions in the UNG gene: 82–304)¹⁸, the bound double-stranded DNA oligonucleotide and 168 crystallographic water molecules. One nucleic acid strand has nine nucleotides (indexed 2–10 with dΨU at 5), whereas the complementary strand has ten (indexed 21–30). Similarities between the crystal structure and the rate-limiting transition state, determined below, suggest that the bound ligand is a transition state analogue. We generated a reactant structure by relaxing the system by energy minimization after interchange of the carbon and nitrogen in positions 1 and 5 of the inhibitor to yield the natural substrate dU. The heavy atom r.m.s. deviation between the minimized and crystal structures was 0.7 Å. The glycosylic bond relaxes from more than 40° to about 25° out of the plane of the pyrimidine base (a superposition of the important protein residues of the X-ray and the minimized structure is shown in Supplementary Information). A natural bond-orbital analysis¹⁹ of model compounds suggests that conformational strain and stereoelectronic effects contribute at most 5 kcal mol^{–1} to catalysis at the calculated angle for the substrate (Supplementary Information).

Two-dimensional adiabatic potential energy surfaces, as functions of the lengths of the bonds that are made and broken (Fig. 1), were calculated by minimizing the energy of the system in the presence of harmonic constraints on those bonds. Figure 1b shows the solvent-corrected energy in terms of the distance between the atoms in the glycosylic bond ($r_{C1'-N1}$) and that between C1' and the oxygen atom of the attacking water molecule ($r_{C1'-OH_2}$). A transition state (TS₁) with energy $E = 14.9$ kcal mol^{–1} (all energies are relative to the reactants) at $r_{C1'-N1} = 2.05$ Å and $r_{C1'-OH_2} = 3.10$ Å (Fig. 2) leads to a well defined intermediate (I₁), with $E = 7.4$ kcal mol^{–1} at $r_{C1'-N1} = 2.75$ Å and $r_{C1'-OH_2} = 2.85$ Å. The presence of the broad minimum at $r_{C1'-N1} = 2.75$ Å and $r_{C1'-OH_2} = 2.85$ Å demonstrates that the mechanism is stepwise and that the glycosylic bond breaks before the water molecule attacks. The leaving group is anionic because there is no effective proton donor in the surrounding region if H268 is neutral, as indicated by NMR⁸. The second transition state in Fig. 1b (TS₂, $E = 10.3$ kcal mol^{–1} at $r_{C1'-N1} = 2.95$ Å and $r_{C1'-OH_2} = 1.95$ Å) corresponds to the attack by the water molecule. The shoulder at $r_{C1'-OH_2} = 1.65$ Å indicates the point at which transfer of a proton to D145 becomes energetically favourable. An analogous

[†] Present address: Department of Chemistry, University of California, Berkeley, California 94720, USA.

scan in $r_{\text{C1}'-\text{OH}_2}$ and the distance between O δ of D145 and the proton in question (data not shown) clearly indicate that earlier ionization of the water would be prohibitively costly in energy (as much as $47.7 \text{ kcal mol}^{-1}$ at $r_{\text{C1}'-\text{OH}_2} = 2.85 \text{ \AA}$). The final products have an energy value equal to $-9.0 \text{ kcal mol}^{-1}$.

Alternative mechanisms of catalysis by UDG are indicated by the yellow and green arrows in Fig. 1b. The green arrows correspond to a perfectly associative pathway in which the C1'–OH₂ bond forms before the C1'–N1 bond breaks; the barrier (truncated in Fig. 1b)

is equal to $94.3 \text{ kcal mol}^{-1}$. A trajectory involving concerted attack of the nucleophile and loss of the leaving group would run parallel to the yellow arrows but its exact placement would depend on the degree to which the mechanism was associative (how close the nucleophile started to C1'). On the basis of these limiting cases, the concerted associative mechanism^{6–8} would fall between the yellow and green arrows and would be markedly higher in energy than the stepwise dissociative mechanism obtained here (red arrows). The fact that the latter mechanism is supported by both the present

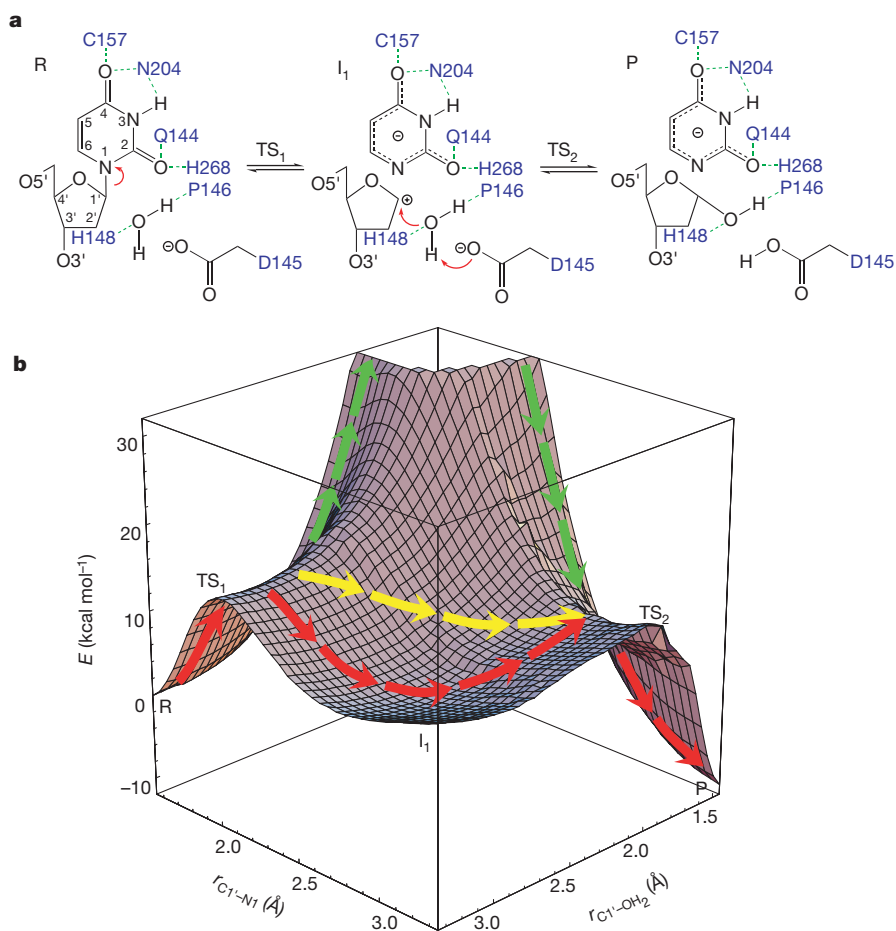


Figure 1 Reaction mechanism. **a**, Schematic diagram (see text). Electron transfers are indicated in red, hydrogen bonds in green and enzyme residues in blue. The dashed line to C157 indicates a $\text{C}\alpha\text{--H}\alpha\cdots\text{O4}$ hydrogen bond³⁰. **b**, Adiabatic potential energy surface as a function of $r_{\text{C1}'-\text{N1}}$ and $r_{\text{C1}'-\text{OH}_2}$. In the region $r_{\text{C1}'-\text{N1}} \leq 2.20 \text{ \AA}$ and $r_{\text{C1}'-\text{OH}_2} \leq 2.00 \text{ \AA}$, the points above 32 kcal mol^{-1} are not shown for clarity. Arrows are drawn by hand to guide

the eye. Red arrows follow the lowest energy pathway (stepwise dissociative); green arrows follow a perfect associative pathway; and yellow arrows follow a concerted associative pathway starting from the reactant structure. The states indicated are reactant (R), product (P), transition states (TS₁ and TS₂) and the oxocorbenium cation/anion intermediate (I₁).

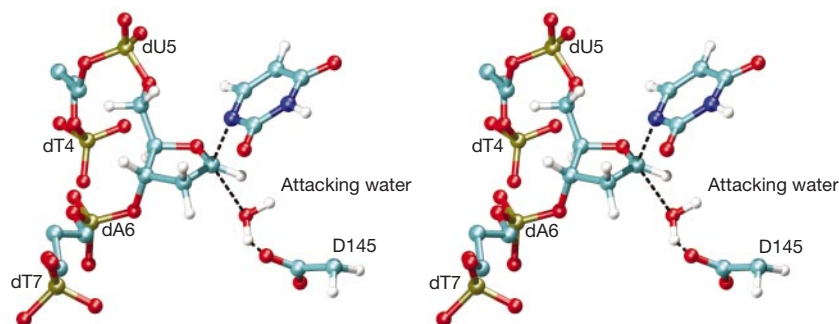


Figure 2 Wall-eyed stereodiagram of TS₁. The dashed lines correspond to $r_{\text{C1}'-\text{N1}} = 2.05 \text{ \AA}$, $r_{\text{C1}'-\text{OH}_2} = 3.10 \text{ \AA}$ and $r_{\text{O}\delta-\text{H}} = 2.01 \text{ \AA}$.

study and recent kinetic isotope experiments¹² is notable as it demonstrates that an oxocarbenium cation intermediate is populated, in contrast to what has been observed in enzyme-catalysed depurination²⁰.

To establish the roles of different groups in catalysis, we cumulatively set to zero the charges of each MM 'residue' (including water molecules and nucleotides) in order of decreasing distance from C1', and recalculated the wavefunction and the solvent corrections—this procedure corresponds to a perturbation calculation in which the structure is kept fixed²¹. The change in the difference in total energy between the reactants and the rate-limiting transition state (TS₁) is plotted in Fig. 3. The groups that result in a rapid rise in the energy as one scans from left to right are those that significantly lower the activation energy of the reaction. The end points give the barrier height along this pathway in the enzyme (14.9 kcal mol⁻¹) and in vacuum (35.7 kcal mol⁻¹). A non-cumulative decomposition is given in Supplementary Information Fig. 2; it is completely consistent with Fig. 3.

The side chains of D145 and H268 (D64 and H187 in *Escherichia coli*) markedly stabilize TS₁ (4.8 and 2.1 kcal mol⁻¹, respectively). These data are consistent with mutation studies in *E. coli*: the D64N and H187Q mutants exhibit 6,600- and 3,500-fold decreases in the maximum single-turnover rates, which correspond to contributions of 5.3 and 4.8 kcal mol⁻¹, respectively⁹. Structurally, H268, which is neutral⁸, donates a hydrogen bond from NeH to O2 of the leaving group. The only other side chain to make a notable contribution is that of H148. This residue, which was taken to be doubly protonated on the basis of continuum electrostatics calculations (pK_a = 9.2, whereas the optimum pH of the recombinant human enzyme is 7.7–8.0 (ref. 18), seems not to have been studied experimentally. It slows the reaction but seems to be important for binding and positioning the reactants as it bridges the 5' phosphate group of dT4 and the attacking water molecule. The side chain of N204 is also expected to be important for specific binding⁶ but does not contribute to lowering the activation energy because it interacts

equally well with the reactants and TS₁. The remainder of the stabilizing interactions in Fig. 3 involve protein-backbone atoms and, most notably, the phosphate groups of the DNA. As for the former, the backbone atoms of C157 stabilize the developing negative charge on O4 (1.8 kcal mol⁻¹), while the carbonyl of P146 interacts with the attacking water molecule (2.5 kcal mol⁻¹); the backbone NH of Q144, which interacts with O2 (refs 6, 7), makes a contribution of 0.4 kcal mol⁻¹. These data suggest that the excess charge on the leaving group resonates to both O2 and O4 (see charges in Supplementary Information), in contrast to recent proposals^{10,22}.

Taken together, the phosphate groups of nucleotides 4–7 make the largest contribution to catalysis (21.9 kcal mol⁻¹). These groups are almost fully buried by the enzyme: their accessible surface area is only 11% of that in B-form DNA. These essentially unshielded negative charges (see Supplementary Information) are positioned in the low dielectric medium of the protein on the side of the sugar ring, opposite that of the nitrogenous base (Fig. 2). The polarization of the reactant (TS₁) structure, including that due to the enzyme, results in a total charge on the leaving group of $-0.37e$ ($-0.76e$) with an opposite charge of equal magnitude on the remainder of the QM deoxyuridine atoms. The negative phosphate charges repel the anionic leaving group, which is farther away in TS₁ and I₁, and stabilize the oxocarbenium cation. Electrostatic stress has also been proposed to promote catalysis in orotidine 5'-monophosphate decarboxylase, but in that case the reactants are destabilized by interactions with specific protein residues²³. In UDG, the large contribution of the phosphate groups, combined with the non-negligible one from protein backbone atoms, accounts for the residual activity of mutant enzymes.

The calculations show that UDG catalyses base excision by a stepwise dissociative mechanism. The glycosylic bond breaks to yield an oxocarbenium cation/anion intermediate. Such a species has not been observed in other enzymes²⁰ but evidence for its presence in UDG is provided by recent kinetic isotope experiments¹². Along this reaction pathway, the highest activation energy (TS₁) is 14.9 kcal mol⁻¹. This number is to be compared with a barrier of 34.5 kcal mol⁻¹ calculated for model compounds in solution, where the nucleophilic attack and loss of the leaving group is concerted (Supplementary Information). The results are in good agreement with measured activation enthalpies (12.1 kcal mol⁻¹ in the enzyme¹⁸ and 32.1 kcal mol⁻¹ in solution²⁴) and account for the 10¹²-fold rate enhancement⁹. Although comparison of TS₁ with the crystal structure¹¹ identifies dΨU as a transition state analogue, it is likely to be the unstrained tautomer with a hydrogen on C1 (Supplementary Information)—stereoelectronic effects and conformational strain¹¹ seem not to be major factors in catalysis. The side chains of the conserved residues D145 and H268 contribute 6.9 kcal mol⁻¹ to lowering the activation energy (counterbalanced by 8.2 kcal mol⁻¹ from the side chain of H148); however, the dominant contribution (21.9 kcal mol⁻¹) comes from the phosphate groups, which are buried on substrate binding, and the protein backbone accounts for an additional 4.7 kcal mol⁻¹. The conclusion that most of the rate acceleration comes from stabilization of an oxocarbenium cation by anionic elements of the substrate is without precedent and could emerge as a general feature of DNA glycosylase enzymes. □

Methods

System studied

The crystal structure of UDG in complex with double-stranded DNA that was used for the calculations is a preliminary version of Protein Data Bank code 1EMH¹¹. We used standard protonation states for all residues other than the histidines. On the basis of NMR data⁸, H268 was taken to be neutral with the proton on Ne. Continuum electrostatics methods²⁵ were used to determine the most probable protonation states of the 12 remaining histidines at neutral pH. The calculations indicated that H148, which forms a bridge between the phosphate group of dT4 and the attacking water molecule, is doubly protonated, whereas the others are neutral with the proton on Ne.

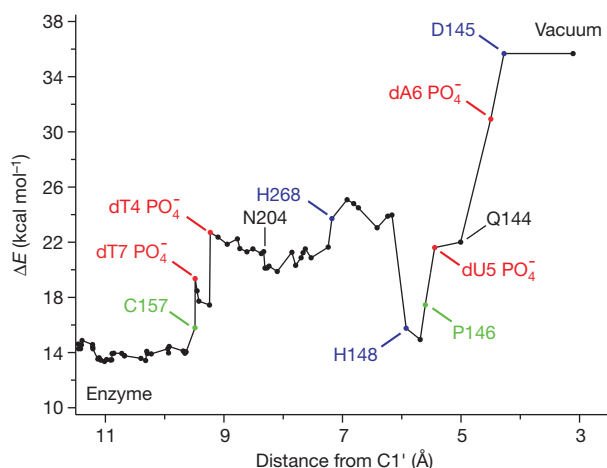


Figure 3 Variation of the TS₁ barrier height relative to reactants as MM charges are set to zero cumulatively in order of decreasing distance from C1' of dU in TS₁. Groups that result in changes of more than 1.5 kcal mol⁻¹ are labelled in colour: DNA, red; protein-backbone interactions, green; and protein side-chain interactions, blue. Additional residues mentioned in the text are labelled in black. Phosphate groups are attached at the 5' position of the nucleotides with which they are indexed and were neutralized separately. To probe its role in this step of the reaction, the side chain of D145 (which does not make a bond in this step) was treated classically with Mulliken charges obtained in the QM/MM analysis. To obtain quantitative agreement with Fig. 1, the solvent corrections were included; they were calculated by repeatedly solving the Poisson equation after setting the charges on each MM group to zero.

QM/MM calculations

We modelled the reaction in the active site with a hybrid QM/MM method⁴. The uracil nitrogenous base, part of the sugar ring (C1', C2', C4', O4' and their associated hydrogen atoms), the attacking water molecule and the side chain of D145 were treated with the semi-empirical method AM1¹⁵. The remainder of the system, which included 167 crystallographic water molecules, was treated classically with the CHARMM all-hydrogen parameter set (that released with version C22 for the protein and with version C27 for the nucleic acid)^{16,27} and the TIP3P water model²⁸. The MM non-bonded interactions were (force) shifted to zero over 8–12 Å, and the QM/MM interactions were truncated at 13 Å. The dielectric constant for the MM electrostatic interactions was $\epsilon = 1$ (but see below).

Because the boundary between the QM and MM regions crosses covalent bonds, it is necessary to add (link) atoms to properly terminate the QM region⁴. There are four such atoms, which correspond to the D145 C α –C β bond, the dU5 C2'–C3' bond, the dU5 C3'–C4' bond and the dU5 C4'–C5' bond. The link atoms were treated as hydrogen atoms that interact with both the other QM atoms and the MM atoms²⁹. Linear geometries and crystallographic bond lengths were enforced for the bonds involving link atoms by additional harmonic terms.

Treatment of solvent shielding

Owing to the fact that the system is highly charged, it is important to include solvent-shielding effects. Explicit treatment of a sufficient number of water molecules to solvate the system fully is not only costly, but is unlikely to yield a realistic dielectric as the solvent cannot reorganize extensively during energy minimization. On the other hand, implicit treatment of the solvent in the vicinity of the reaction is also inappropriate as the lengths and directionalities of specific hydrogen bonds can have important roles. Consequently, we adapted a procedure introduced for free-energy perturbation calculations¹⁷. In step I of the procedure, the charges of ionic groups (all 8 arginines, all 18 lysines, all 12 glutamates, the 8 aspartates other than D145, and all 17 phosphate groups) were scaled. We chose the scale factors according to the potential-based scheme described in ref. 17. We then minimized the system with the scaled charges in the presence of a limited number of explicit solvent molecules to obtain a reasonable set of structures. In step II, the energy associated with restoring the scaled charges to their full strengths in vacuum was calculated, and in step III, the system was transferred from vacuum ($\epsilon = 1$) to bulk solvent ($\epsilon = 80$). In steps II and III, a grid-based method was used to solve the Poisson equation and the QM atoms were represented by their Mulliken charges—the procedure differed slightly from that outlined in ref. 17 to account correctly for changes in the MM atom positions. The solvent-corrected effective energies reported are the sums of the energies obtained in step I and the electrostatic solvation free energies obtained in steps II and III.

Higher-level QM calculations

To check the AM1 results, we performed simulations with higher-level quantum methods. Energies were calculated with both Hartree–Fock (HF) and density functional theory (B3LYP) with a 6-31++G** basis set for the system described above at the geometries of the extrema in Fig. 1b. In addition, the later parts of the reaction (attack by the water molecule and the proton transfer) were refined with further minimizations in which the leaving group was treated as MM with the AM1 Mulliken charges for I₁ (see Supplementary Information), and the remainder of the QM atoms were treated by HF with a 4-31G basis set. HF/6-31++G** and B3LYP/6-31++G** energies were calculated for these structures as well. It was found that the agreement for the first part of the reaction was very good but that AM1 introduced a small artifactual barrier to the proton transfer, which is necessary to complete the reaction. The physical basis for this discrepancy is that, at larger values of $r_{\text{C1}'-\text{OH}_2}$, AM1 favours a structure in which the water molecule forms two nonlinear hydrogen bonds with O δ of D145, whereas HF leads to one hydrogen bond with O δ of D145 and another with the carbonyl group of P146. The barrier in the AM1 calculations derives from the need to reorient to an HF-like geometry as $r_{\text{C1}'-\text{OH}_2}$ decreases.

Received 15 May; accepted 22 August 2001.

- Wood, R. D., Mitchell, M., Sgouros, J. & Lindahl, T. Human DNA repair genes. *Science* **291**, 1284–1289 (2001).
- Mol, C. D., Parikh, S. S., Putnam, C. D., Lo, T. P. & Tainer, J. A. DNA repair mechanisms for the recognition and removal of damaged DNA bases. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 101–128 (1999).
- Nilsen, H. *et al.* Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol. Cell* **5**, 1059–1065 (2000).
- Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comp. Chem.* **11**, 700–733 (1990).
- Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. The structural basis of specific base excision repair by uracil-DNA glycosylase. *Nature* **373**, 487–493 (1995).

- Slupphaug, G. *et al.* A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92 (1996).
- Shroyer, M. J. N., Bennett, S. E., Putnam, C. D., Tainer, J. A. & Mosbaugh, D. W. Mutation of an active site residue in *Escherichia coli* uracil DNA glycosylase: effect on DNA binding, uracil inhibition and catalysis. *Biochemistry* **38**, 4834–4845 (1999).
- Drohat, A. C. *et al.* Heteronuclear NMR and crystallographic studies of wild-type and H187Q *Escherichia coli* uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **38**, 11876–11886 (1999).
- Drohat, A. C., Jagadeesh, J., Ferguson, E. & Stivers, J. T. Role of electrophilic and general base catalysis in the mechanism of *Escherichia coli* uracil DNA glycosylase. *Biochemistry* **38**, 11866–11875 (1999).
- Drohat, A. C. & Stivers, J. T. NMR evidence for an unusually low N1 pK_a for uracil bound to uracil DNA glycosylase: implications for catalysis. *J. Am. Chem. Soc.* **122**, 1840–1841 (2000).
- Parikh, S. S. *et al.* Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl Acad. Sci. USA* **97**, 5083–5088 (2000).
- Werner, R. M. & Stivers, J. T. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: evidence for an oxocarbenium-ion anion intermediate. *Biochemistry* **39**, 14054–14064 (2000).
- Haushalter, K. A., Stukenberg, P. T., Kirschner, M. W. & Verdine, G. L. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.* **9**, 174–185 (1999).
- Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **460**, 165–181 (2000).
- Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
- Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217 (1983).
- Simonson, T., Archontis, G. & Karplus, M. Continuum treatment of long-range interactions in free energy calculations. Application to protein-ligand binding. *J. Phys. Chem. B* **101**, 8349–8362 (1997).
- Slupphaug, G. *et al.* Properties of a recombinant human uracil-DNA glycosylase from the *UNG* gene and evidence that *UNG* encodes the major uracil-DNA glycosylase. *Biochemistry* **34**, 128–138 (1995).
- Weinhold, F. In *Encyclopedia of Computational Chemistry* (ed. Schleyer, P. R.) 1792–1811 (Wiley, New York, 1998).
- Chen, X.-Y., Berti, P. J. & Schramm, V. L. Transition-state analysis for depurination of DNA by ricin A-chain. *J. Am. Chem. Soc.* **122**, 6527–6534 (2000).
- Bash, P. A. *et al.* Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* **30**, 5826–5832 (1991).
- Drohat, A. C. & Stivers, J. T. *Escherichia coli* uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry* **39**, 11865–11875 (2000).
- Wu, N., Mo, Y., Gao, J. & Pai, E. F. Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase. *Proc. Natl Acad. Sci. USA* **97**, 2017–2022 (2000).
- Shapiro, R. & Kang, S. Uncatalyzed hydrolysis of deoxyuridine, thymidine, and 5-bromodeoxyuridine. *Biochemistry* **8**, 1806–1810 (1969).
- Schaefer, M., Sommer, M. & Karplus, M. pH-dependence of protein stability: absolute electrostatic free energy differences between conformations. *J. Phys. Chem. B* **101**, 1663–1683 (1997).
- MacKerell, A. D. Jr, Wiórkiewicz-Kuczer, J. & Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **117**, 11946–11975 (1995).
- MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- Reuter, N., Dejaegere, A., Maigret, B. & Karplus, M. Frontier bonds in QM/MM methods: a comparison of different approaches. *J. Phys. Chem. A* **104**, 1720–1735 (2000).
- Desiraju, G. R. The C–H...O hydrogen bond: structural implications and supramolecular design. *Acc. Chem. Res.* **29**, 441–449 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank N. H. Williams for a critical reading of the manuscript; X. Lopez, Q. Cui and R. Petrella for helpful discussions; B. Webb for technical assistance; and W. G. Richards for his hospitality in Oxford, where most of the work was carried out. A.R.D. is a Burroughs Wellcome Fund Hitchings-Elion Postdoctoral Fellow, and, when the research was initiated, M.K. was the Eastman Visiting Professor in Oxford. The work at Sheffield is supported by a grant from the Biotechnology and Biological Science Research Council; the work at Harvard is supported in part by a grant from the National Institutes of Health.

Correspondence and requests for materials should be addressed to M.K. (e-mail: marci@tammy.harvard.edu).

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayouni (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Welss (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (6F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager:

Kate Cowan

naturejobs

nature.com

Right physics problem, but...

The Institute of Physics this month published a report (www.physicsweb.org) warning of a shortage in the number of UK-trained physicists. Although the overall number of undergraduates taking physics degrees has remained the same, more of them are being siphoned off into fields such as engineering and information technology.

The contributing factors to the problem are diverse — but they start in education. Far fewer UK physics students are choosing to become teachers. In the early 1990s, almost 600 of the roughly 2,400 new physics graduates went into teaching. Now that figure has dropped to about 200. Meanwhile, a quarter of existing physics teachers are over 50 years old, compared with just 11% who are under 30.

The Institute of Physics has a plan to combat the predicted shortage. It recommends a new kind of degree that would accept students with poorer maths skills, then attempt to improve those skills during courses that would expose the students to an interdisciplinary approach to problem solving.

It's a creative response to a vexing issue, but one can't help but wonder whether it really addresses the root of the problem — especially in the context of other data within the same report. For example, the report mentions decreased support for physics infrastructure, perhaps most evident in the "chronic under-funding" of laboratory infrastructure, which has led to the loss of over 10 university departments over the past decade.

It's likely that undergraduates who would make good candidates for existing physics degrees are aware of this lack of support, and choose to pursue other fields. And it's also possible that students who lack the mathematical savvy of their peers still have the ability to know a poor career outlook when they see one.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

New head for UK biology research council; an astrophysicist goes home; biotech reshuffles; and more

[Back page](#)

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

GOVERNMENT

Julia Goodfellow, vice-master and head of the School of Crystallography at Birkbeck College, University of London, will in January become the first female chief executive of the UK Biotechnology and Biological Sciences Research Council (BBSRC). Goodfellow will control annual research grants worth more than £250 million (US\$370 million). She anticipates that soon more women will take top positions in the life sciences, reflecting the number who have taken up biology during the past few decades. "It's evolutionary and natural that this will happen," she says. But she's concerned that women aren't making similar advances in the physical sciences — the discipline in which she originally trained. Besides distributing the BBSRC's funds, Goodfellow sees attracting more British students into the sciences as a top priority. "In the United States, a lot of young people in science are not US nationals," she notes. She doesn't want that to be the case in Britain.

PHYSICS

Astrophysicist **Jeremiah Ostriker** has left Princeton for the University of Cambridge. The former Princeton provost was appointed Plumian professor of astronomy and experimental philosophy last month. The move is, in a way, a return to his roots: Ostriker was a postdoctoral fellow at Cambridge. Afterwards, he joined the Princeton faculty, was promoted to full professor in 1971 and to the Young professorship in 1982. From 1979 to 1995 he chaired the department of astrophysical sciences and served as director of the Princeton University Observatory, and in 1995 he was appointed provost. Ostriker, who last autumn was awarded the US national medal of science, will continue to hold his Princeton post as Charles A. Young professor of astronomy.

Michael Allen became director of the University of Warwick's new Centre for Scientific Computing this month, as well as professor of physics at the UK university. Previously, Allen was professor of physics at the University of Bristol, where he spent 16 years of his career. His research emphasis is on what he calls "molecular virtual reality" — conducting computer simulations of liquids and liquid crystals, using a range of models to investigate the link between the structure of molecules and their properties in different phases. The new centre, with its commitment to large-scale computing resources, provides a good fit for his interests. "The centre provides an exciting opportunity to bring together computational scientists from a range of disciplines, encouraging

them to interact and share methods and ideas," Allen says. He sees his role shifting from focusing on his own work to fostering multidisciplinary collaborations. The centre aims to add 10 permanent positions during the next five years.

BIOTECH

Patrik Dahlén last month left a high-level position with instrumentation company PerkinElmer to take a top position in biotechnology. Dahlén was president of the life-sciences division at PerkinElmer, and is now chief executive of BioImage, a Copenhagen-based company that was spun off from Novo Nordisk in 1999. After Dahlén's departure, PerkinElmer reshuffled its own management, appointing John Engel executive vice-president and president of life sciences.

Rick Dyer this month left his position as head of chemical synthesis at Roche Discovery in Welwyn, UK, to become director of research and non-clinical development with Ardana Bioscience, an Edinburgh-based company focusing on reproductive health. At Roche, Dyer helped to identify 10 clinical candidates and bring three to market. Before joining Roche, Dyer spent four years at Chiroscience in Cambridge, England, where he helped build the company's finances from £3 million in venture-capital funding to a £45-million valuation on the London Stock Exchange. Dyer gained his first degree at the University of Manchester and his PhD in chemistry at the University of Southampton. He then undertook a two-year post-doctoral fellowship at Harvard University. After that, Dyer returned to Britain and spent six years as a senior research chemist at Glaxo in Ware.

UNIVERSITY ADMINISTRATION

Regis Kelly, professor and former chair of the department of biochemistry and biophysics at the University of California, San Francisco (UCSF), assumed the university's executive vice-chancellorship this month. A cell biologist and neuroscientist, Kelly has been a member of the UCSF faculty for more than 30 years, also serving as director of the Hormone Research Institute and chair of biochemistry and biophysics. Kelly now oversees a \$350-million annual research programme. He plans to increase the university's research presence in stem-cell research, human genetics and bioinformatics. He replaces Zach Hall, who plans to join EnVivo Pharmaceuticals, a new biotechnology company located in the Bay Area, as president and chief executive.



Julia Goodfellow



Michael Allen



Patrik Dahlén

CONTACTING US AT MOVERS

Please send any information on new appointments or job moves to: naturejobseditor@naturedc.com.